

ABSTRACT

Title of Dissertation: Identifying and Eliminating the Problem with Einstein's Cosmological Constant

Zachary Myers, Doctor of Philosophy, 2005

Dissertation directed by: Professor Jeffrey Bub – Department of Philosophy
and

Professor James Mattingly – Department of Philosophy

The cosmological constant, Λ , was first introduced into Einstein's field equations in the early 20th century. It was introduced as a quantity of outward-pushing energy in space that would counteract the contracting force of gravity thereby keeping the universe in a balanced and static state. Einstein willingly removed it once the universe was observed to be dynamic rather than static. However, as the decades have gone by, Λ has maintained its supporters and has continually been reintroduced to solve problems in cosmology. Presently, there is good reason to believe that Λ or something like it is indeed present in our universe. In the 1960s, in an effort to provide a physical basis for Λ , particle physicists turned to quantum vacuum energy and have since estimated a value for Λ to be $\sim 10^{110} \text{ erg/cm}^3$, which happens to be significantly greater than its observationally constrained value of $\sim 10^{-10} \text{ erg/cm}^3$. This discrepancy of 120 orders of magnitude has come to be known as the cosmological constant problem. Any effort to resolve the inconsistency must also account for the various observations we attribute to Λ , such as cosmic inflation and cosmic acceleration. To date, there are two basic approaches to

resolving the cosmological constant problem that we may call the Identity approach and the Eliminativist approach. The Identity approach entails that vacuum energy is responsible for all the relevant observations and the problem is to be solved by some cancellation mechanism within the internal components of the vacuum. The Eliminativist approach explicitly rejects the reality and cosmological efficacy of vacuum energy, seeks alternative explanations for the observations and eliminates the cosmological constant problem by eliminating the cosmological constant. The benefit of having a crisis between these two views at this particular stage in cosmology's history is that they can be tested against each other in an experimental situation. Whatever the outcome of the experiments, we will be clearer about the work needed to resolve the cosmological constant problem once and for all.

IDENTIFYING AND ELIMINATING THE PROBLEM WITH
EINSTEIN'S COSMOLOGICAL CONSTANT

By

Zachary Myers

Dissertation Submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:

Professor Jeffrey Bub, Chair
Dr James Mattingly
Dr Mathias Frisch
Dr Ramin Sina
Dr Don Perlis

© Copyright by

Zachary Myers

2005

Acknowledgements

Many people have helped me write this dissertation. My greatest debt goes to my advisor James Mattingly for helping me pin down various points made in this work, as well as his great suggestions and his attentiveness to my concerns. He has been vital in my developing a better sense of the steps necessary to do truly effective research in philosophy. I would also like to thank Jeffrey Bub for being my other advisor throughout the writing of this work and for his tranquil and accommodating demeanor while assisting me through the various other stages of my PhD program. I'd like to thank Mathias Frisch for the insight he lent me during our discussions in his office and for agreeing to be on my committee. Thanks also to Ramin Sina, for agreeing to be a member of the committee as well as for the countless chalkboard sessions we had when we worked together at the Institute for Physical Science and Technology. The writing of this work was made easier because of my experience doing physics research in the past. To this extent I offer much gratitude to Dr Eun-Suk Seo for giving me the chance to do some truly novel research in high-energy astrophysics. From those years of working for her I learned first hand what kind of a commitment is required to make substantial advances in the fields relating to cosmology. I'd like to also thank Professor Don Perlis for agreeing to serve as the Dean's Representative. And a final acknowledgement to all those personally close to me who took enough of an interest in what I do to ask about it.

Table of Contents

<u>Section 1: Introduction</u>	1
<u>Section 2: Cosmology as an Experimental Science</u>	10
2.1: On Observational and Experimental Science	12
2.2: From Old Astronomy to Modern Cosmology	28
2.3: Cosmology as Historical Science and the Testing of Historical Hypotheses	31
<u>Section 3: The History of Λ</u>	40
3.1: The Cosmological History of Λ	40
<i>Early days of Λ</i>	43
<i>Developments against Λ</i>	46
<i>Incentives for maintaining Λ</i>	52
3.2: The Quantum History of Λ	60
<i>Early Quantum History</i>	60
<i>Recent Quantum History</i>	67
<u>Section 4: The Development of the Cosmological Constant Problem</u>	71
4.1: The rise of the Cosmological Constant problem	71
<i>Further Developments of Vacuum Energy</i>	71
<i>Development of the Standard Model</i>	73
4.2: Interpreting the Cosmological Constant Problem	76
4.3: Recent Observations and Relevant Cosmological Facts	83
4.4: The Current Status of the Cosmological Constant and the Problem	89
<u>Section 5: The Identity Approach to Solving the Cosmological Constant Problem</u>	94
5.1: Supersymmetric Solution	96
5.2: String theoretic Solution	102
5.3: The Semiclassical Solution	108
<i>Assuming the Validity of the Semiclassical Approach</i>	109
<i>Assuming the Reality of Quantum Vacuum Energy</i>	118
<u>Section 6: The Eliminativist Approach to Solving the Cosmological Constant Problem</u>	123
6.1: The Essence of Quintessence	125
<i>The rise of dark energy</i>	125
<i>Tracker models of Quintessence</i>	130
6.2: Rethinking Inflation	134
<i>On Inflation</i>	134
<i>Original Conception</i>	135
<i>Current Status of Inflation</i>	136
<u>Section 7: The Experimental Approach to the Cosmological Constant Problem</u>	140

<u>Section 8: Summary and Conclusion</u>	149
<u>Appendix: The Anthropic Approach to Solving the Cosmological Constant Problem</u>	154
<u>On the anthropic principle</u>	155
<u>Anthropic Solutions</u>	157
<i>Weinberg's view</i>	157
<i>Vilenkin's view</i>	159
<u>Bibliography</u>	163

Section 1: Introduction

Since its debut appearance in Einstein's field equations (1917), the cosmological constant, Λ , has had something of a controversial status. Originally introduced as a problem solver, it has seemingly created just as many along the way. Einstein's primary motivation for adding the cosmological term to his gravitational field equations was to create a model of the cosmos consistent with his theory of general relativity (GR) that was static as well as homogeneous and isotropic – features that Einstein then thought described the large-scale structure of the actual cosmos. Einstein officially abandoned the cosmological constant after Hubble's redshift observations indicated that the universe is not static but expanding (Einstein, 1931). These observations were accommodated by Friedmann's (1922) expanding universe solutions to Einstein's field equations without a cosmological constant.

Even though Einstein abandoned it, some members of his community did not then nor ever follow suit for various reasons. In *"Lambda: The Constant that Refuses to Die"*, Earman says, "Although the cosmological term has from time to time, been perceived to give off a bad odor, it has stubbornly refused to go away – partly because (Einstein notwithstanding) there is a logical place for it in GTR, partly because it has always had its champions, and partly because it gets periodically called upon to solve problems and resolve crises in cosmology," (2001, p. 189).

As has often been the case, these problems get resolved shortly afterwards in other ways, which then removes the need to invoke Λ , and a movement emerges to send

Λ back to the shelf until a new problem arises. This wavering trend has given Λ a rather dubious status in physics up to the present day.

This cosmological parameter has been reconceptualized in various ways, leading to further complications as to where it fits into the subject of current cosmology. In the late 1960s, efforts were made by particle physicists to provide a physical basis for Λ , rather than settling for Einstein's *ad hoc* move of putting it into his field equations by hand to ensure a static universe. As a result of these efforts from the particle physicists, an inconsistency arose that has come to be known as the cosmological constant problem. In essence, quantum field theory (QFT) predicts that 'zero-point' fluctuations occur in the lowest energy state or vacuum state of a quantum system. Contrary to a classically defined vacuum as empty space, the quantum vacuum state is dynamic and contains energy. This feature had been understood for years and it was thought that the cosmological constant could arise from the fluctuations that occur within a ground state (sometimes called the vacuum state). In a given quantum vacuum, these fluctuations can accumulate to eventually create an enormously high energy density state, and such a high energy density state is what modern cosmologists have come to believe the universe was in at its very beginning. Moreover, each of the individual terms that build up the vacuum energy density (represented as ρ_{vac} or ρ_{gs})¹ is calculated to be much larger than the value allowed by current observational constraints. This inconsistency between quantum field theoretic calculations and observation is the cosmological constant problem. From GR, the theory used in cosmology, the observational value for Λ is close to zero and when

¹ It should be noted here, with an apology from the author for the chosen notation, that when ' ρ ' or ' ρ_0 ' appears in equations it refers to the matter-energy density of the whole universe, whereas ρ_{gs} or ρ_{vac} refers strictly to quantum ground state energy density, which may be a contributor to the total ρ_0 . And ρ_{gs} and ρ_{vac} will be used interchangeably in this work.

interpreted in terms of vacuum energy density, is $\sim 10^{-8} - 10^{-9} \text{ erg/cm}^3$. Whereas QFT has predicted that Λ is *significantly* greater. When one considers the state of the very early universe, quantum theoretical effects become non-negligible. At the Planck era, $\sim 10^{-43}$ seconds after the big bang, the vacuum energy density is calculated to be $\sim 10^{110} \text{ erg/cm}^3$. This discrepancy of ~ 120 orders of magnitude has become a problem for both cosmologists and particle physicists alike.

The cosmological constant problem has also become a noteworthy problem for those with an interest in the growing fields of history and philosophy of scientific cosmology as well. Even though cosmology has been part of human thought since the beginning of our history, there has been limited philosophical inquiry into scientific cosmology. This lack of attention may be due to the fact that for a large part of that history, cosmology was not widely regarded as a 'real' science. The status of cosmology for much of its history resembled myth more than science. Stories of the cosmos, its origin and nature, were just as likely to show up in religious texts as in the writings of natural philosophers. Obviously for much of our history we lacked the technological means for making various observations that could have supported claims about the cosmos. However, in recent years there have been significant observational and theoretical advances in the subject. Cosmology has developed to a point where hypotheses can be made and tested with observations. Cosmology has become a mature science allowing for experimental situations, and through experimentation cosmologists have the means of possibly solving the cosmological constant problem or at least to point them in the right direction of a solution. At this stage some people (scientists and others)

may not be convinced that cosmology is a science susceptible to rigorous experimentation. This issue will be addressed in section 2.

In section 2 we will thoroughly examine the role experimentation has played in science. From its inception during the scientific revolution to the present day advances in science have come from a variety of experimental methodologies. The history of science does not suggest one definitive way in which experimentation in science is supposed to work. At times hypotheses have preceded observation and at other times observations have preceded hypotheses, as well as several instances when theory and observation have developed independently of each other and fit together beautifully. From these instances we may appropriately define experimental situations as observational situations that have come to bear upon some particular hypothesis *in the appropriate way*, meaning that the set of observations must be able to support counterfactual claims of a theory. Observations that do not link up to a hypothesis in this way are not experiments.

We will further see that the long-standing distinction between the observational sciences on the one hand and the experimental sciences on the other is neither a helpful nor accurate way to classify the sciences. What this means for cosmology is that what was once considered an observational science at best is actually a science that allows for experimental situations to occur within its domain as effectively as any other science. At this stage in cosmology's development, the ability to conduct experiments is more useful than ever, especially for the debates surrounding the cosmological constant problem. Toward the end of the thesis we will see more precisely how experimentation can work to help resolve the problem. But first we will have to review the problem itself in some detail to get a sense of what an adequate solution would entail.

The cosmological constant problem is not a problem for QFT or GR in isolation, but emerges when these two theories are considered together. The advocates of both GR and QFT have postulated what they take the value of the cosmological constant, Λ , to be. In order to put the cosmological constant problem into a clearer light it is best to give a historical summary of the critical events, discoveries, and research programs that were occurring in early 20th century physics. This will be done in section 3. Since the cosmological constant problem stems from two theories, the history of Λ could be thought of on two fronts, or at least consisting of two periods, which I shall designate as: The Cosmological History of Λ , and the Quantum History of Λ . The cosmological history precedes the quantum history of Λ . My account of the former will focus on the events that led up to Einstein's introduction of Λ and its ensuing problems, followed by some further incentives that cosmologists had for maintaining Λ . This narrative will also indicate the critical points in cosmology's recent history that led to its experimental development. The latter begins with the conception and discovery of quantum vacuum energy, its development in the field of quantum physics and eventually interpreted as a cosmological constant that can explain certain cosmological observations such as the age of the universe, distance of quasars, and various other effects attributed to inflation.

In discussing the conceptual origins of the cosmological constant, we can begin to see the implications of the cosmological constant problem for methodological issues in the philosophy of science and for foundations of physics. Section 4 will focus more on the cosmological constant problem itself. Historically, the cosmological constant problem has gone through various stages as a problem. The progression of the changes began when Zel'dovich (1967), a particle physicist, realized in the late 1960s that if these zero-

point fluctuations are occurring and $\Lambda \neq 0$ then it was a serious issue because the vacuum energy may have gravitational effects that cannot be ignored. In an attempt to derive Λ from physical constants, a discrepancy emerged with observational constraints between 8 and 40+ orders of magnitude. Later, in the early 1970s as a result of theoretical physicists extrapolating even further back in the early universe with the emergence of electroweak unification and spontaneous symmetry breaking the discrepancy grew to ~ 120 orders of magnitude. Then with the advent of inflationary cosmology in the early 1980s, a large vacuum energy density was a required feature to provide the cosmos with the initial boost to instigate inflationary expansion. The problem went from explaining a numerical inconsistency to explaining how a large numerical value vanished completely after inflation. In the 1990s new observations were made that significantly reshaped our conception of the cosmos regarding its structure and dynamics. These observations indicate that roughly 70% of the cosmos is filled with an unknown or ‘dark’ energy that is causing it to accelerate. Some cosmologists have called upon Λ (or vacuum energy) as the dark energy candidate to account for these observations, which *explicitly* suggests $\Lambda \neq 0$. The problem went from explaining how Λ vanished completely after inflation to vanishing almost completely but leaving a little bit behind to produce what we have observed.

Originally the cosmological constant problem was an epistemic problem in that there must be some mechanism that is responsible for the vanishing of Λ and we just have not figured out all the details yet. If $\Lambda \approx 0$, then this should be accounted for by some sort of cancellation mechanism between the individual contributions to the QFT vacuum energy as Λ decreases significantly during the phase transitions in the early

universe. But since the introduction of inflationary cosmology in the 1980s and the recent observations in the 1990s, what has been important is that this canceling mechanism must cause Λ to vanish during a series of symmetry breakings, and not suggest that Λ was *always* miniscule, otherwise some other source would have to be suggested for driving inflation. As we evaluate various cancellation schemes we will see this as an ongoing issue, (but it does not mean that alternative inflationary schemes do not exist).

The cosmological constant problem could also be characterized as an ontological problem. The difficulty of explaining how Λ has vanished stems from the difficulty in reconciling the ontological status of Λ in QFT with its status in GR, and that no solution can be properly posed until we have a better sense of the ‘cosmic stuff’ that Λ allegedly represents, (experimentation is helpful to this end). Others have considered the cosmological constant problem to be a pseudo-problem because it involves doing QFT in a curved spacetime, and to this extent it cannot even be posed in a physically meaningful way. There is good reason to believe that it is physically meaningful despite these concerns, but is nevertheless a pseudo-problem in that it only needs to be taken seriously if one believes that Λ should be derivable from other fundamental constants in physics. The idea of a possible relation between Λ and fundamental constants in particle physics is appealing for it would give some indication that a grand unified theory of physics is possible. But, at present, we cannot say if this is possible. By evaluating the problem of the cosmological constant in these various respects, it will put us in a better position to evaluate the solutions posed to it. Sections 5 and 6 will be primarily devoted to reviewing such solutions. In these sections we will revisit the significance of experimental situations in cosmology.

Any approach to solving the cosmological constant problem must be able to give an explanation for the numerical inconsistency while at the same time accounting for all the relevant observations. The observations that would need to be explained could be roughly categorized as early and current, that is accounting for the observed effects of inflation such as spatial flatness and uniformity (from the early universe) and the observed density and acceleration (in the current universe). The solution types put forth toward the problem can be categorized into two basic views: the Identity view and the Eliminativist view.

Since the inception of the problem in the late 1960s until the observations made in the 1990s much of the work done to solve the problem had been trying to find some cancellation mechanism to explain how all the vacuum energy disappeared, and if such a mechanism could be adequately described, the problem would be solved. This approach presupposed that the cosmological constant was vacuum energy, $\Lambda \equiv \rho_{gs}$, and so solutions were thought to be found by making theoretical reconstructions often involving QFT in some way. Solutions were sought through supersymmetry, super gravity, then eventually Superstring theory. Efforts to dissolve the problem with a semiclassical approach to QFT were also sought. Some of these attempts were more successful than others, but they were and still are, at best, promissory notes. The problem until then was a problem of resolving a large numerical inconsistency – vacuum energy was not thought to have any *current* cosmological significance. However, once the recent observations were made, these Identity theorists designated Λ (or quantum ground state energy) as the unknown ‘dark’ energy responsible for all of these observed effects. The Eliminativist approach is not committed to such a claim. In fact, the Eliminativist approach explicitly rejects the

existence of vacuum energy and so is committed to either rejecting all phenomena that vacuum energy is responsible for (either hypothetical or observed) or providing alternative explanations for them. While the Eliminativists need not doubt the recent observations, they attribute another cause to them; instead of the cosmological constant, the observations may be attributable to another dark energy candidate such as ‘Quintessence’. If something else is responsible for the universe’s early inflationary period as well as its 70% missing content and acceleration, then the cosmological constant could indeed be zero, and accounting for the 120-order inconsistency (between the early universe and today) could be disregarded as an irrelevant puzzle. In essence, to eliminate the cosmological constant is to eliminate the cosmological constant *problem*.

The fact that these two approaches, Identity and Eliminativist, have emerged from the recent observations is significant because it now allows physicists to take an experimental approach to the cosmological constant problem. When we ask the questions of ‘what is the universe made of?’ and ‘what’s causing it to accelerate?’, these can be answered experimentally. Both the Identity theorists and the Eliminativists have different answers regarding the energy source that is responsible; their dark energy candidates can be tested against one another experimentally. The outcome of such experimentation would be of great assistance to physicists concerned about the cosmological constant problem, for it would give them a more precise direction to field their research. If experimentation can be done to rule out vacuum energy as a dark energy candidate, then Λ and all its inconsistencies can be swept aside. If experimentation leads to evidence in favor of vacuum energy as the most probable dark energy candidate, then physicists will be much clearer about the work that lies before them for telling a consistent story of Λ .

Section 2: Cosmology as an Experimental Science

The history of science over the past few hundred years suggests that experimentation has taken on several roles. History indicates that theories and the data that support them can be related in various ways. Great advances in science have come from well-conducted observations that have supported a theory, and likewise great theories have emerged from very astute observations. Some of the greatest advances in science, and especially cosmology, have even come from unexpected yet happy family get-togethers of theory and observation where each were made for different reasons but meshed wonderfully together. If the object is to learn about nature, we do this by effectively relating our ideas of nature (hypotheses) with our observations of it. Observation need not strictly precede theory nor theory guide observation. I claim that the temporal ordering is not particularly relevant; what is relevant is *how* our observations come to bear on our hypotheses. Provided a set of observations come to bear upon a hypothesis *in the appropriate way*, we can say that an experiment has taken place. What this means is that the set of observations must be able to support counterfactual claims of a theory. Thus, *to do an experiment is to make counterfactual-supporting observations or indicate how such observations can support the counterfactual claims of a theory*. Observations that do not link up to a hypothesis in this way are not experiments. To say that experimentation aids in the establishment (or demise) of a theory, we mean that by making reliable counterfactual-supporting observations our theories become established (or perish).

When we say that counterfactual support is necessary for experimentation, we tend to think that tweaking or manipulating the system under observation is necessary.

The ability to intervene with a system has, to date, been the central criterion that distinguishes the ‘experimental’ sciences from the ‘observational’ sciences. The observational sciences such as astronomy, astrophysics, planetary science, and various elements of marine biology, as well as the historical sciences (which are a subset of the observational sciences) such as geology, paleontology, and evolutionary biology, often focus on features of nature that do not permit intervention in the literal or conventional sense. We cannot tweak the conditions of, say, the sun, stars, or various other cosmological entities nor can we manipulate the events that occurred in the distant past. Since intervening enables us to determine the causally relevant factors that are responsible for what we see, some scientists tend to think that without this ability, we don’t have access to the causes of what we observe, and so the situation is strictly observational. However, I claim it is incorrect to suggest that the same level of rigorous experimentation cannot be done in the domains of the observational (or historical) sciences. Hypotheses about a particular phenomenon or system can gain counterfactual support even if the system in question does not permit us to intervene directly with it, but understanding how this is done will require a reevaluation of what constitutes an intervention. In the following subsections we’ll see how cosmology (which has been classified as both an observational and historical science) allows for experimentation as effectively as any classic experimental science. By the end of this section it will become clear that there’s no need to classify the sciences as being either observational or experimental, but only to classify observational or experimental *situations* within particular sciences.

2.1: On Observational and Experimental Science

There has been a tradition to divide the sciences up into the experimental on one side and the observational on the other. Generally, when we speak of an “experimental” science it is thought to be one where hypotheses are tested in some controlled setting. When scientists speak of an “observational” science it is often depicted as one where its practitioners go out into the field and observe nature directly. The distinction of labels is clear enough, and many scientists accept this segregation. Paleontologist, T. Wolosz has said that “in Experimental Science we can conduct a direct test or experiment to verify our hypothesis ... in observational science we cannot experimentally test our hypothesis).”² This distinction exists in the sciences related to cosmology as well. As quoted from the reviews of COSMOLOGY: The Big Bang's Radical Brother, (2000), W. Hogg and M. Zaldarriaga say, “In an observational science like cosmology, theories are not subject to laboratory experiments but are evaluated by their explanatory relations.”

The label “observational” seems to imply that only mere observations of nature are being made, as though the objective of observational scientists is solely to describe or classify what they see, and that no real testing of scientific hypotheses is really occurring. If this is the case, then according to W. Salmon, “a ‘science’ that consisted of no more than a mere summary of results of direct observation would not deserve the name,” (Salmon, 1967, pg. 5). The distinction between the observational and experimental sciences, as noted in the quotes above, suggests that something special is going on over in the experimental sciences that the observational scientists cannot participate in, and that

² Dr. Tom Wolosz at the Center for Earth & Environmental Sciences, SUNY at Plattsburgh (2005): <http://faculty.plattsburgh.edu/thomas.wolosz/>

all these scientists can do is make observations but not do any experiments themselves. But is this an accurate way to classify scientific disciplines? It is true that we cannot do any direct experimenting on any dinosaurs today and we cannot put the entire cosmos under a microscope, but this does not mean that the sciences of paleontology, astronomy and cosmology do not permit experimentation. If the observations made in a particular science can be brought to bear on some hypothesis and produce counterfactual support, one has all one need's to conduct an experiment. If the scientists in these fields can formulate hypotheses and make these kinds of observations, then why wouldn't experimentation be possible? Keeping in mind what was mentioned in the previous subsection, the specific order in which the hypothesis and the observations are made in not the critical part, what is critical is if they can relate to each other in the right way.

C.E. Cleland has noted that even though the practices of doing field research in the 'observational' sciences are different from classic experiments in a controlled laboratory setting, as we shall see, these sciences that have generally been labeled as observational similarly allow for experimental situations. Because each practice is tailored to exploit the information that nature puts at its disposal, and because the character of that information differs, neither practice may be held up as more objective or rational than the other. "Aware that experiment plays different roles in science, skeptical about the existence of a single method for all of science, and unable to provide an epistemically satisfying account of the rationality and objectivity of any scientific practice, [some] philosophers have been reluctant to generalize, let alone make normative judgments, across different disciplines," (Cleland, 2002, pg. 475).

What classifies a situation as experimental has much to do with the practice of making observations under well-specified conditions that primarily support counterfactual claims. Counterfactual support is critical for distinguishing experimental and observational situations in science, for if an experimental situation is one where observations are testing some law-like claim, the hypothesis would need to be tested under varying conditions to see if it holds. Observational situations are those where observations or measurements are made strictly for gathering information rather than to test any hypothetical claim (such as gathering samples of specimens or mapping the night sky). Experimental situations must provide evidence to support certain counterfactuals.

In order to characterize the practice of science more precisely, some version of the observational/experimental distinction may be worth preserving, but modified as a classification scheme within particular sciences rather than between them. No particular science need be thought of as strictly experimental or strictly observational. It seems incorrect to assume that practitioners of any particular scientific discipline spend all of their time testing hypotheses in a laboratory or all of their time making observations out in the field, as if no overlap between these two types of science exists. Given how science is done in reality, many of the sciences allow for both experimental and observational situations within their discipline. For instance, often times scientists go out into the field and simply collect data: making measurements of certain natural structures such as rock formations, water temperatures, etc. Or they go out and collect samples of fossils, specimens, etc. These data are often collected without any specific hypothesis in mind that they are intending to test. However, what is important is that these data *could* be used to test a hypothesis. In many instances of observation, just because there are no

hypotheses for the data to support does not mean that relevant hypotheses could be not formulated, nor does it mean that the data could not be related to or come to support counterfactual statements.

For example, in 1827 a botanist, Robert Brown, reported on the irregular movement of pollen suspended under water. This Brownian motion had been observed by others even 60 years before; some thought it was vital action of living pollen itself. Brown made painstaking observations but for a long time it came to nothing. However, in the first decade of the 20th century there was simultaneous work being done by experimenters such as J. Perrin and theoreticians such as Einstein, who showed that pollen was being bounced around by molecules. These results lent persuasive empirical support to the kinetic theory of gases and finally converted even the greatest skeptics, (as appearing in Hacking, 1983).

As mentioned, learning about nature is done by relating our ideas of it with our observations of it. The temporal ordering of theory and observation is less important than ensuring that the observations come to bear upon the theory in such a way that they could support counterfactual claims. In the case of Brownian motion above, Brown's data supported counterfactual claims that could be made about the kinetic theory of gases, such as 'if the temperature and pressure of fluids that we detect on the macro level is the result of scores of molecules bumping around on the micro level, then when small enough objects such as grains of pollen are placed into that fluid, there should be a noticeable displacement effect due to these molecules colliding with it.' Even though as a matter of historical fact the theory and observations were made at different times and for different reasons, this case has all the necessary features of an experimental situation. To

make the distinction clear, in an observational situation one collects data, in an experimental situation one employs that data to support or falsify a hypothesis. One more example made to disambiguate observational and experimental situations from the early 19th century is worth noting here.

Claude Bernard's work, Introduction to the Study of Experimental Medicine (1865) is a classic attempt to distinguish the concepts of experiment and observation. He challenges the observational/experimental classification with many different examples from medicine where the distinction between observation and experiment gets murky. He cites an example of a doctor who, in the war of 1812, was in the position to observe the workings of the digestive tract of a man with a terrible stomach wound over an extended period of time. Does this constitute an experiment or was it just a sequence of observations under unique circumstances? Based on the description, all we can tell is that a sequence of observations were made; we don't know anything about the intentions of the doctor making them. However, given the criteria for experimentation that have been presented so far, we can say that these observations could have been used in an experiment. They could even be part of a current experiment if someone were to formulate a hypothesis entailing the relationship between, say, certain symptoms of indigestion and a stomach wound of this nature. For instance, "no one with a punctured pancreas will be able to completely digest high-starch foods within the first week of the puncture wound." The observations made of this single wounded soldier may not be sufficient on their own to confirm the hypothesis, but could contribute to its confirmation if part of a larger data set involving various subjects with a similar type of wound.³ Also,

³ Human experimentation often requires special safeguards against outside variables such as the placebo effect. Such human experiments can be conducted in a 'double-blind' way, meaning that neither the

gaining support for this hypothesis would involve making observations under varied conditions in order to make further related claims, such as ‘If one has a stomach wound of this nature, digestion of high-starch foods is virtually impossible during the first 3 days.’ Or ‘If one has a stomach wound of this nature, the digestion of high-starch foods becomes possible after around 9 days.’ Or ‘If one has a stomach wound of this nature, digestion of high-starch foods becomes possible after 3 days if one takes this kind of medicine.’ When we consider experimental situations as those observational situations that have been brought to bear upon a specific hypothesis in a way that offers counterfactual support, then the ambiguity between experiment and observation goes away.

Some scientists would believe that this medical case could have indeed constituted an experiment because of our ability to intervene with the system. Administering medicine to injured people and taking note of its effects would constitute an intervention, however, hypotheses about a patient’s recovery rate could still be formulated and tested even if no intervention of this kind occurred. Still, many scientists, such as the one’s quoted above who endorse the observational/experimental distinction, maintain that without the ability to intervene experiments cannot be done. This may be why many scientists tend to think that experimental science is done in the lab or some other controlled setting and for those who focus their research on systems where literal intervention is not possible, such as in astronomy, geology, and cosmology, all that can be done is raw observation. By this reasoning, observations made about a system where literal interventions cannot be made do not count as experiments even if they lend

volunteer nor the researcher knows which individuals are in the experimental ‘test’ group until after all the data are collected. The double-blind control feature ensures that any effects on the volunteer are due to the treatment itself and are not a response to the knowledge that he/she is being treated.

empirical support to a hypothesis. I think this is incorrect. The same level of rigorous experimentation can be done in the domains of the ‘observational’ sciences. Once a theory is in place experiments can be done to test its predictions whether in a controlled laboratory setting or not. Hypotheses in the form of counterfactual claims can still be tested through experimentation and we can still speak meaningfully about causal relations even if the system in question does not permit us to intervene directly with it. Understanding how this is done requires a reevaluation of what it means ‘to intervene’. Let us now consider how we may come to speak about causal relations in the events we observe.

By *literally* intervening in a physical system and seeing the changes made we can gain access to the causal knowledge necessary to test our claims. The basic idea in the recent literature on causation is that of a “surgical” change in *A* that leads to a change in *B* occurs only as a result of its causal connection, if any, to *A* and not in any other way. In other words, the change in *B*, if any, that is produced by the manipulation of *A* should be produced only via a causal route that goes through *A*. Manipulations or changes in the value of a variable that have the right sort of surgical features have come to be called *interventions* in the recent literature⁴.

These manipulability theories of causation, according to which causes are to be regarded as devices for manipulating effects, have considerable appeal and have been popular among certain philosophers as well as social scientists and statisticians, (see Collingwood (1940), Gasking (1955), von Wright (1971), and Menzies and Price (1993)). von Wright (1971) describes the basic idea of the manipulability approach as follows,

⁴ From Jim Woodward’s article in the Stanford encyclopedia for philosophy:
<http://plato.stanford.edu/entries/causation-mani>.

“... to think of a relation between events as causal is to think of it under the aspect of (possible) action. ... if p is a (sufficient) cause of q , then if I could produce p I could bring about q ,” (p. 74). A very similar view has been put forth by P. Menzies and H. Price (1993). Both von Wright and Price and Menzies seem to embrace the further conclusion that follows from this, namely, human action must be a concept that is more ‘basic’ than the notion of causality. By appealing to the notion of agency to provide a reductive analysis of causation, these authors claim that we have a grasp of the *experience* of agency that is independent of our grasp of the general notion of causation.

In their paper ‘*Causation as a Secondary Quality*’ (1993) Price and Menzies advance an “agency” theory of causation. Their basic thesis is that “an event A is a cause of a distinct event B just in case bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B , (1993, pg. 187)⁵. For Price and Menzies, causation is a secondary quality – something that the agent perceives and attributes to nature, but it would be misleading to claim that causality exists *in* nature. Nature can provide the data, the sequences of observable events, the seemingly lawlike regularities but to speak of the *causal* relations between the events is an epistemic issue. For philosophers like von Wright, Price and Menzies, causality is epistemic not metaphysical. Nowhere in nature do we actually see causes, we only see lawlike regularities. From our experiences as agents we may witness a sequence of events and

⁵ They take this connection between free agency and causation to support a probabilistic analysis of causation (according to which “ A causes B ” can be plausibly identified with “ A raises the probability of B ”) provided that the probabilities appealed to are what they call “agent probabilities,” where agent probabilities are to be thought of as conditional probabilities, assessed from the agent’s perspective. (1993, p. 190)

note that often times when A occurs, B occurs. We describe this as ‘A causes B’. For instance, from the observations such as, ‘*when the pot of water comes in contact with the stove element, the temperature of the water increases*’, and ‘*the longer we keep the pot of water in contact with the stove element the greater its temperature increases*’, we tend to say that the pot of water coming in contact with the stove element causes its temperature to increase. Speaking this way is how agents relate events in the world, but according to Price and Menzies, there are no intrinsically *causal* properties in the stove element itself. Similarly, agents can speak about seeing red things in the world, but ‘red’, like all colors, is a secondary quality.

James Woodward takes issue with these agency accounts of causality. Firstly, according to Woodward, designating agency as primary, and reducing all instances of causation to instances of human agency is circular. After all, what is agency if not a form of causation? For Woodward, causes are real, they exist in nature, and the reason why we witness various lawlike regularities is precisely because of these causal relations. For Woodward, by speaking about causality as something that we attribute to a system once an agent has intervened and seen how the appropriate changes occur is problematic. Surely, by intervening in a particular system, and seeing that by altering A there is an alteration in B, is sufficient to speak meaningfully about A causing B. However, for Woodward, his issue is regarding the causal relations of systems where literal interventions are simply not possible, such as the causal relationship between the gravitational attraction of the moon and the motion of the tides, or the the causal relationship between drought and a crop’s yield. If the only way to understand causation is by means of our prior grasp of an independent notion of agency, then it is hard to see

what could justify us in extending the notion of causation to circumstances in which manipulation by human beings is not possible and the relevant experience of agency is not available. Woodward (2000) and J. Pearl (2000) claim to have circumvented these problems by reformulating the manipulability approach in terms of the notion of an intervention that is characterized in purely causal terms that make no essential reference to human action. Some human actions will qualify as interventions but they will do so in virtue of their causal characteristics, not because they are carried out by humans, (see footnote 4).

According to Woodward a process or event can qualify as an intervention even if it does not involve human action or intention at any point. He characterizes an intervention entirely in terms of causal and correlational concepts and makes no reference to human beings or their activities. For Woodward, a purely “natural” process involving no animate beings at all can qualify as an intervention as long as it has the right sort of causal history. Moreover, even when manipulations are carried out by human beings, it is the causal features of those manipulations and not the fact that they are carried out by human beings that matters for recognizing and characterizing causal relationships. For example,

“under this approach X will qualify as a (total) cause of Y as long as it is true that for some value of X that if X were to be changed to that value by a process having the right sort of causal characteristics, the value of Y would change. Obviously, this claim can be true even if human beings lack the power to manipulate X or even in a world in which human beings do not or could not exist. There is nothing ... that commits us to the view that all causal claims are in some way dependent for their truth on the existence of

human beings or involve a “projection” on to the world of our experience of agency,” (see footnote 4).

Woodward critiques these philosophers, von Wright, Price, and Menzies, who analyze causality away in terms of human agency, but they need not be as worried about speaking causally where literal intervention is not possible as Woodward suggests. Woodward takes their view to imply that if agency and literal intervention is not possible then we cannot speak meaningfully about causes. For these philosophers, observations can still be made where we see lawlike regularities, and we can still speak meaningfully about causes, but they are simply not committed to any claims about the perceived objects possessing any intrinsic causal properties. For these philosophers, to put human agency as the primary feature of their theories of causality is not to imply that without literal intervention no causal talk can be done, but only that causes come from the human agents that attribute them to the natural systems in question rather than being embedded in nature as Woodward believes. In actuality, despite Woodward’s worries, these philosophers would be able to accept an account of interventions as he suggests, and would be able to accept the possibility of making observations of a particular system that support counterfactual claims even when literal intervention is not possible.

von Wright, Menzies and Price recognize that there are obviously some situations in nature where literal intervention is not possible and have responded accordingly. Menzies and Price claim that we can gain knowledge of the causal relations between events where intervention is not possible if these events are similar to a set of events where manipulation *is* possible. In other words, we can come to know the cause of a particular token of a certain type that we cannot manipulate provided we have gained

causal knowledge of a similar token of the same type through an intervention. They argue that:

“when an agent can bring about one event as a means to bringing about another, this is true in virtue of certain basic intrinsic features of the situation involved, these features being essentially non-causal though not necessarily physical in character. Accordingly, when we are presented with another situation involving a pair of events which resembles the given situation with respect to its intrinsic features, we infer that the pair of events are causally related even though they may not be manipulable ... Clearly, the agency account, so weakened, allows us to make causal claims about unmanipulable events such as the claim that the 1989 San Francisco earthquake was caused by friction between continental plates. We can make such causal claims because we believe that there is another situation that models the circumstances surrounding the earthquake in the essential respects and does support a means-end relation between an appropriate pair of events. The paradigm example of such a situation would be that created by seismologists in their artificial simulations of the movement of continental plates,” (1993, p. 197).

The response of Menzies and Price is similar to von Wright’s earlier view that to understand a causal claim involving a situation that human beings cannot in fact manipulate such as ‘the eruption of Mt Vesuvius caused the destruction of Pompeii’, we must interpret it in terms of claims about causal situations that human beings *can* manipulate. Whether one wants to think of causes metaphysically, as things that are really part of the world, as Woodward does, or to interpret causes epistemically, attributed to objects and events in the world by agents as a way of presenting the world in

terms we can understand, it is nevertheless clear that we can speak meaningfully about causal relations even when literal intervention is not possible.

By accepting the legitimacy of these kinds of (non-literal) interventions, an intervention can tell us what should be envisioned as changed and what should be held fixed when we evaluate a counterfactual like ‘If X were to be changed by an intervention to such and such a value, the value of Y would change’. By interpreting counterfactuals in this way we can make sense of causal claims in contexts where human interventions do not in fact occur and arguably even in cases in which they are causally impossible, as long as we have some principled basis for answers to questions about what *would* happen to the value of some variable *if* an intervention were to occur on another variable.

Hence, experimental situations that support counterfactual claims can still be done once we consider interventions in this way. If we accept this interpretation of intervention, and we accept interventions as necessary features of experimental situations, then experiments can be done in nature, even in its farthest reaches, as effectively as in a controlled setting. Members of the observational sciences such as geology, paleontology, marine biology, astronomy, and astrophysics, and cosmology have the means of doing experiments. Hypothesis testing can still be done with *natural* experimentation. An example of a natural experiment would be Eddington’s eclipse experiment to confirm General Relativity.

According to GR, stars near the sun would appear to have been slightly shifted because their light has been curved by the gravitation field. Stated differently, if spacetime is as GR says it is, the stars near the sun would appear to have been slightly shifted because their light has been curved by the gravitation field. This effect could only be noticeable during an eclipse, since otherwise the sun's brightness obscures the stars. The question was whether this effect could be observed. In 1919 during an eclipse, Arthur Eddington took pictures of the stars in regions around the sun that were consistent with this prediction. His observations lent empirical support to the theory⁶.

The ability to do natural experiments is often enhanced by nature's tendency to be fruitful and to multiply; in essence, she can often times provide us with many positive instances of the particular thing we're scrutinizing. In such cases, the body of evidence available to researchers may resemble what an experimental program would provide, and they may exploit this fact by formulating generalizations analogous to those formulated by classic experimentalists. If we wished to extend the techniques of Eddington's experiments, we could say that the more massive the object, the more it curves the space around it, the more the light ray will bend, and the more shifted the stars behind the massive object will appear. Even though we cannot literally intervene with a system and, say, increase the mass of a galaxy, and then see the increased shift in the position of stars behind it due to the increased curvature of space, we can observe a variety of other instances of more massive galaxies in the cosmos and see the enhanced displacement of

⁶ Even though these results were hailed as a conclusive proof of GR over Newtonian gravity, recent historical examination indicates that Eddington was arbitrarily selective with the results he used.

the stars behind them. The cosmos provides many positive instances of this. Observing the shifted position of stars behind massive objects such as galaxies and galaxy clusters is called gravitational lensing and has become a very effective technique in modern astronomy. In fact, gravitational lensing has become so well established because this kind of experimentation is possible.

We may consider counterfactuals as being reasonably supported by the data if we interpret the data as what we would observe if we did intervene, or we interpreted the data as obtained from a set of events that is similar in type to some other set of events where we could intervene. Hence (causal) claims about cosmological phenomena can be empirically supported by considering interventions as being logically possible even if not actually possible.

A point to be noted is that when intervention is not possible *even in principle*, no causal knowledge can be gained about a particular event or system. When a system permits no interventions whatsoever, no counterfactual support can be obtained. For instance, particle physics tells us that the nuclei of certain atoms are prone to decay and if we were to arrange a handful of say, Uranium atoms in exactly the same state, our theories of particle physics suggest that some of them will decay and some will not. For the Uranium atoms that do decay, we could never know the cause of this because not only are we unable to literally intervene but there's no circumstance under which we could construct a token instance that would support a counterfactual statement like 'were we to manipulate the system in this particular way then the nucleus would have decayed sooner rather than later'. We cannot speak meaningfully about intervening in situations that we cannot even in principle create. When we cannot speak meaningfully about

interventions we cannot speak meaningfully about causal relations between events. And so, whether you take causes to be primary qualities of things (such as Woodward) or secondary qualities of things (such as Price and Menzies), it remains the case that where no type of intervention whatsoever is possible, then no counterfactual-supporting observation is possible, hence, no experimentation is possible.

What we have seen is that interventions need not be thought of as being strictly human interventions. The varying conditions of a particular system in question that occur as the result of intervention can be provided by nature itself. If we were to accept this account of intervention and counterfactual support, then ‘intervention’ could no longer function as a distinguishing criterion between observational and experimental science, since it could occur in both traditions.

By reevaluating interventions and counterfactual support in the above-mentioned way, it is now clear that experiments can be done in the domains that scientists once called merely observational. The sciences of astronomy, astrophysics, and cosmology can ground experiments and test hypotheses as effectively as any scientific research done in a controlled laboratory setting. To this extent, we should note that not only is it possible to do rigorous experimentation in current cosmology, but by this account of experimentation it has already been going on for some time now. The results of these natural experiments have sculpted our conception of the cosmos over time. The impact of experimentation can indeed be seen and felt in our current conception of the universe and has in turn affected the experimental motives of current cosmologists, as we shall now see.

2.2: From Old Astronomy to Modern Cosmology

Some quick definitions and distinctions may be helpful here. The differences between the subjects of astronomy, astrophysics, and cosmology are not always clear, and certainly there is much overlap between them as well. Astronomy has often been thought of as the definitive observational science. For thousands of years, the task of astronomers, who often worked under the patronage of monarchs or noblemen, was to make observations of the heavens. Astronomy existed long before the scientific revolution and hypothesis testing, and for much of astronomy's history it really was only an observational activity. But in modern times astronomers take a much more scientific approach toward their subject. Modern astronomy takes an in depth look at the contents of the heavens rather than just mapping the distribution and movement of stars across the sky. The development of modern astronomy is certainly attributable to technological advances over the past few hundred years and astronomers can now study and analyze the matter of outer space, especially the dimensions, composition and evolution of celestial bodies. In order to make any progress beyond what was discovered in ancient times, modern astronomy requires the knowledge and application of a certain amount of physics. Astrophysics, then, is the study of the physics of the universe. When using physics to explain various features of the heavens, one is doing astrophysics. Hence, this may be a semantic issue, but much of the 'astronomy' done from Newton's time up to the present could be more precisely described as astrophysics. And cosmology could be characterized as the science of theoretical astrophysics on the largest scale that allows for experimental situations. Cosmology is most simply characterized as the study of the universe as an entity. That entails the study of its structure and dynamics as a whole,

rather than its parts, as well as its origin, history and evolution. These distinctions between astronomy, astrophysics, and cosmology, as suggested, allow for significant overlap. Broadly speaking, any inquiry directed toward the heavens and outer space is essentially a cosmological inquiry. Any astronomical observations that are brought to bear on a cosmological hypothesis in a way that supports counterfactuals is a cosmological experiment.

It is by appropriately defining an experimental situation as one where a series of counterfactual-supporting observations are brought to bear on a hypothesis for empirical support or refutation, that we may remove any ‘observational vs. experimental’ distinction between the sciences and legitimately classify cosmology as a science that allows for experimental situations. The history of cosmology indicates that, while allowing for experimental situations, it has not progressed in accordance with one specific experimental method. That is to say, progress through experimentation did not come solely from some Baconian inductive approach, nor from some strictly hypothetico-deductive approach. In fact some of the greatest leaps in cosmology came from the unexpected but happy meetings of theory and observation (as mentioned earlier). The two most salient examples are how Hubble’s observations of galaxy recession (1929) supported Lemaitre’s budding theory of the Big Bang (1927). And how Penzias and Wilson’s unexpected and perplexing discovery of cosmic microwave background radiation (CMBR, 1965) related to the then slightly more established theory of the big bang. In both of these cases, the scientists were not specifically motivated to obtain empirical support for the theory that their data eventually supported. Nevertheless, a cosmological hypothesis regarding the origin of the universe found empirical support

with their results, and their data from galaxy recession and CMBR helped the Big Bang hypothesis win out over its competitors, namely Einstein's static model of the cosmos and Hoyle's steady-state model, respectively. These events will be discussed in more detail in the next section.

Some points to be noted are that even though the history of cosmology suggests that great progress has come from these family get-togethers of theory and observation, rather than some premeditated scientific hypothesis that motivated a highly successful experiment, what we see is that cosmology allows for all the right things to fit together for experimentation to occur. A further point about these happy-family experimental situations is that they not only helped to further our understanding of the cosmos but also incited us to reformulate our conception of it. Over the past several hundred years our conception of the cosmos went from being originally geocentric to being heliocentric, then eventually to doing away with heliocentricity all together; then from being static to being dynamic, and from being eternal and infinite to being finite in both space and time.

At present, given the substantial empirical support that has accumulated in favor of the Big Bang hypothesis, it has become virtually the standard conception in contemporary cosmology. Our current conception of the cosmos as one that originated from a point-like explosion some 14 billion years ago has significant consequences for what we hope to discover about it. Before the big bang theory of the cosmos emerged counterfactual-supporting observations were made to test hypotheses about the universe's structure, contents, and dynamics, but this was always against the backdrop of a universe that was eternal and infinite. Questions about the universe's origin, its history and evolution were simply never asked, at least not in a scientific context. Now, in light of

our current big bang model, cosmology has acquired an entirely new and *historical* dimension, and so formulating and testing these historical hypotheses adds another dimension to the experimental situations in this domain.

2.3: Cosmology as a Historical Science and the testing of Historical Hypotheses

As mentioned in the beginning of this section, the historical sciences can be thought of as a class within the observational sciences. Many of the sciences that have been described as ‘observational’ have a historical element to them, such as astronomy and cosmology, as well as paleontology and geology. As was shown that the observational sciences permit experimental situations in their own right, we’ll now see how observations in these specific ‘observational’ sciences come to bear on hypotheses to obtain counterfactual support. The focus of the research in the historical sciences is on explaining existing natural phenomena in terms of long past causes. Such work is significantly different from making a prediction and then artificially creating a phenomenon in a laboratory, or looking to nature for a variety of positive instances of the phenomenon in question. Good historical research involves the formulation of a causal theory that can unify various (seemingly disparate) phenomena, then looking for the appropriate evidence to support the theory by supporting its counterfactual claims.

Some would say that testing these historical hypotheses does not actually amount to real experimentation. For instance, Henry Gee, an editor of the science journal *Nature* has remarked that the hypotheses formulated in the historical sciences “can never be tested by experiment, and so they are unscientific. . . . No science can ever be historical” (2000, 5–8). In other words, for Gee, a genuine test of a hypothesis requires classic

experimentation in some controlled setting. Under this view, paleontology, evolutionary biology, geology and cosmology are not only non-experimental sciences, they're not sciences at all! Once again, this is because some would believe that we don't have access to the causally relevant events of the past that are responsible for what we observe. But this is simply not true in many cases.

Some historical research does also allow for a process analogous to observing regularities among event-types. For instance, as has been commonly stated, given the finite speed at which light travels, to look up at a star system is to look into the past. So, if one were to hypothesize about how a particular star system emerged, one could execute a natural experiment of observing various other stars systems similar to the one in question at different stages in their development. Again, due to nature's tendency to repeat herself, there are plenty of star systems up in the sky that could support counterfactual claims like 'were the cloud of hydrogen to be less massive, a star system like the one in question would not have emerged' or 'if the volume of hydrogen were dispersed beyond such and such a volume, no star system could form'. Similarly, given the abundant evidence of fossil remains, no one doubts the hypothesis that dinosaurs once existed, that they had the propensity to grow to enormous sizes relative to humans, and that some of them were herbivores and others were carnivores. Likewise with the abundance of remaining fossils providing convincing evidence that humans evolved from an ape-like ancestor many people nowadays do not doubt the theory of evolution.

When nature is generous and provides us with an abundance of tokens from the past there is little doubt that the body of evidence used to support a hypothesis can closely resemble that provided by classic experimental research. But what if nature is not

so generous; do natural experiments cease to be possible? No. When nature is not so generous and provides only a single or few instances of a particular phenomenon, such as, say the Big Bang, historical researchers have to appeal to another one of nature's tendencies – her tendency to leave traces of past events. For historical researchers, finding a trace or 'smoking gun' of an event is the pivotal evidence necessary for gaining observational support for a hypothesis and/or for discriminating between competing hypotheses. Smoking guns provide evidence for hypotheses that retrodict past events just as successful, empirically adequate predictions provide evidence for the generalizations examined in the lab. Instead of inferring test implications from a target hypothesis and performing a series of controlled experiments, historical scientists focus their attention on formulating differing causal etiologies for the traces they observe, (see Cleland, 2002). These causal explanations are then presented as mutually exclusive hypotheses and the search continues for evidentiary traces to discriminate among them. Because of the asymmetry of overdetermination, there are often a large number of subcollections of the effects of a past event that are individually sufficient (given the right theoretical assumptions) to infer its occurrence⁷. The trick is finding the right ones to discriminate between rival hypotheses, (Cleland, 2002, pg. 488).

We can see that the evidential reasoning of the historical sciences is notably different than that employed in the kinds of experiments mentioned thus far. Smoking

⁷ According to David Lewis (1991, 65–67) "the asymmetry of overdetermination" is a basic feature of the world. The basic idea is that events leave widespread, diverse effects. Any one of a large number of traces is sufficient (given the laws of nature) to conclude that the event occurred. One doesn't need every shard of glass in order to infer that a window broke. A surprisingly small number of appropriately dispersed fragments will do. The hard part of historical research is finding traces that are capable of unambiguously discriminating among rival hypotheses. But even when much time has passed and the traces have become very attenuated there is frequently the possibility of discovering a smoking gun for a long past event.

guns are obviously not produced by systematically varying some conditions while holding others constant such as in a controlled setting. But similar to the natural experiments described above, they are uncovered in the uncontrolled world of nature by fieldwork. Attempting to test hypotheses when the data are scarce should not be seen as a strike against this kind of experimentation for two reasons.

One is that there are plenty of instances in ‘classic’ experimental contexts where the control over critical parts of a system is very limited, repetition of events is difficult, and data are scarce. A prime example of such experiments is the search for high-energy particles from collisions in an accelerator. In these cases, experimenters often hope to find a single golden event; much in the same way as a historical researcher searching for a smoking gun. The other reason is that in these situations when the natural data are scarce, hypotheses can nevertheless eventually obtain counterfactual support by considering other hypotheses that follow from them. The Pangea hypothesis from geology is a good example of this.

By the early 20th century there was already some evidence that all the continents on Earth were once all part of a large single land mass, Pangea. There were the complementary shapes of the Atlantic coasts of Africa and South America, as well as similarities in geological formations and fossil records on opposite sides of the Atlantic Ocean. Thus, some time in the past pieces of this giant continent may have drifted apart somehow and created the individual continents we see today. In 1912, Alfred Wegener first advanced the theory of continental drift to support the Pangea hypothesis. However it was not widely accepted; partly because of the scarcity of evidence and partly because there was no known causal mechanism for horizontal continental motion. The competing

hypothesis at that time, the contractionist theory, held that the Earth's crust moves only vertically, as a result of gradual contraction as its interior cools.

In 1960 Princeton geologist Harry Hess advanced the further hypothesis of sea floor spreading to support the continental drift hypothesis. Hess's hypothesis was formulated in light of new discoveries made in geophysical oceanography; these discoveries included the global mid-oceanic ridge system, the apparent youth of the sea floor, and the circum-Pacific island arc-trench system with its numerous volcanoes and powerful earthquakes, (Cleland, 2002, pg. 481). At roughly the same time ideas were being advanced that the very mechanism that causes the Earth to generate a magnetic field (by the convection of molten iron and nickel) also causes it to reverse its magnetic polarity after intervals of millions of years. The subsequent discovery of alternating bands of reversed magnetism, spreading out symmetrically on both sides of the volcanically active Mid-Atlantic Ridge provided the smoking gun for sea floor spreading. These magnetic stripes provided compelling evidence that the Earth's crust was moving horizontally as Wegener had suggested, carrying the continents along with it. The observation was consistent with counterfactual claims that if the Earth's crust moved horizontally and the Earth's magnetic field had switched directions several times over the past billions of years, then we would expect to see this kind of distribution on the sea-floor. Furthermore, the observations of a young ocean floor with magnetic stripes etched into it were more consistent with Wegener's continental drift hypothesis than any counterfactual claims regarding the vertical movement of the Earth's crust from the contractionist model. The empirical support for continental drift in turn supported the primary hypothesis of Pangea.

The evolution of Wegener's hypothesis (from continental drift to sea floor spreading to plate tectonics) to support the Pangea hypothesis underscores a notable feature of successful historical work. As Philip Kitcher (1989, pgs 430–432) and others have emphasized, the best scientific explanations are unifying. In successful historical science the focus is on the unity that a hypothesis provides for a diverse body of puzzling traces. Successful historical hypotheses explain traces by *unifying* them under a consistent *causal* story. The plate-tectonic and Pangea hypotheses came to be accepted not just because of the observations that supported the counterfactual claims derived from them, but also because of their ability to unify ostensibly independent phenomena.

Geologists often refer to the development of plate tectonics as a revolution in their field analogous to the development of the periodic table in chemistry and Darwin's theory of evolution through natural selection. The Pangea case parallels the developments of Big Bang cosmology in the 20th century quite well. The big bang model successfully unified various phenomena, such as the relative abundance of light elements over heavy elements, galaxy recession, and CMBR, by designating them all as effects of one major event in the early universe. The big bang model was able to unify these phenomena better than any competing static model of the cosmos. It was the final discovery of CMBR as a distinct remnant of a great cosmic explosion that persuaded cosmologists and physicists alike to accept the Big Bang theory.

At the time, there also existed the Steady State theory as an alternative to the Big Bang and both of these theories could account for Hubble's data on Galaxy recession. However, the Big Bang was better able to account for CMBR as the trace of a major cosmic explosion more easily than its steady state competitor. As mentioned above, the

experimental intentions of Penzias and Wilson, the discoverers of CMBR, were very different from what their data actually came to support, but this does not deter from the relevance of their work. The discovery of CMBR and its cosmological significance was also an instance of an unexpected but happy meeting between theory and observation.

As the story goes, in the early days of transatlantic radio there was always a lot of static. Many sources of the noise could be identified, although they could not always be removed. By the 1930s Karl Jansky at Bell Labs had located a ‘hiss’ coming from the center of the Milky Way. Hence, there were sources of radio energy in space that contributed to the familiar static, (Hacking, 1983, pg. 158).

In 1965, the radio astronomers A. Penzias and R.W. Wilson adapted a radio-telescope to study this phenomenon. They expected to detect energy sources and that they did, but they also found a small amount of energy that seemed to be uniformly distributed everywhere in space. It would be as if everything in space that was not an energy source were about 3 degrees Kelvin. Since this did not make much sense to them, they did their best to discover instrumental errors. As the often-cited example, they thought that some of this radiation may have come from the pigeons that were nesting in their telescope, and they painstakingly had to remove them and all their remains. But after they had eliminated every possible source of noise, they were still left with their uniform 3 degrees. They were hesitant to publish because a completely homogeneous background radiation seemed literally nonsensical.

Fortunately, just as they had become certain of the meaninglessness of the phenomenon, Robert Dicke and his team of Princeton physicists were circulating a preprint that suggested, in a qualitative (and ostensibly counterfactual) way, that if the

universe had begun with a big bang, there would be a uniform temperature throughout space; the residual temperature of the first explosion. Moreover they predicted that this residual energy would be detected in the form of radio signals. The observational work of Penzias and Wilson meshed perfectly with this hypothesis.

Penzias and Wilson had shown that the temperature of the universe is almost everywhere about 3 degrees above absolute zero and this residual energy was taken as the smoking gun of creation. It was a compelling piece of evidence and the first truly compelling reason to believe in the Big Bang over the steady state theory. It seemed reasonable to believe (counterfactually) that if the universe had begun with a cosmic explosion then we would expect to see some residual energy. The story of how CMBR came to bear upon the big bang theory is a prime example of how historical hypotheses can gain empirical support, how observations and theory can successfully come together and lead to progress in cosmology, and how cosmology in particular can allow for experimental situations by providing the right kinds of observations that can support the counterfactual claims of a theory.

In summary, what we have seen through these various examples of observation and theory is that these two features of science can relate to each other in various ways, experimentation can be done for various reasons, and can be done to support various kinds of hypotheses. Theory need not precede observation; after all, Penzias and Wilson went exploring and found the uniform background radiation prior to any theory of it. But once a theory is in place that makes predictions it can be tested and experiments can be conducted whether in a controlled environment or not. Once a set of observations in any scientific domain can be brought to bear upon a relevant hypothesis in a way that

supports counterfactuals, an experimental situation can arise. These events can occur in any science and because of this, there is no need to distinguish between observational and experimental *sciences*, but only to distinguish between observational and experimental *situations* in science. In light of these modifications to the definition of experiment, cosmology need not be thought of as a purely theoretical science laden with speculative claims that are far beyond our technological means to test. New phenomena and the competing hypotheses to account for them can be tested in cosmology just as effectively as any experimental situation that could arise in any other science.

As experimentation has helped cosmology in the past to progress this far, it gives us hope of resolving problems in current cosmology. In particular it may be through experimentation that progress can be made with the problems surrounding Λ . As it stands, Λ has been brought in to account for recent cosmological observations, namely to account for the current density and acceleration of the cosmos. These recent observations and Λ 's relation to them have contributed to the already complicated cosmological constant problem. However, alternatives to Λ exist to account for these observations. The debate between these competing hypotheses could be settled with a set of experiments. If it does not yield a particular result that resolves the cosmological constant problem entirely, then at least it would rule out other possibilities and will provide cosmologists with a more precise direction to field their research. Let us now turn our attention to the cosmological constant.

Section 3: History of Λ

Section 3.1: The Cosmological History of Λ

In GR, gravity is no longer a force as it was in Newton's theory of gravity but is a consequence of the curvature of spacetime. The theory provides the foundation for the study of cosmology and has given scientists the tools for understanding many features of the universe that were not discovered until well after Einstein's death. Mathematically, Einstein modeled space-time as a four-dimensional manifold and his field equations state that the manifold's curvature at a point is directly related to how the matter and energy are distributed at that point. In other words, matter distribution and the curvature of spacetime are intricately related.⁸ Einstein's original field equations were written as follows:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = (8\pi G/c^4)T_{\mu\nu} \quad (1)$$

where $R_{\mu\nu}$ is the Ricci curvature tensor, R is the scalar curvature, $g_{\mu\nu}$ is the metric tensor, $T_{\mu\nu}$ is the stress-energy tensor (the term that represents the matter-energy distributions and gravitational fields), c is the speed of light and G is the gravitational constant, which also occurs in Newton's law of gravity. The Einstein equations relate the geometry of space in the form of the Ricci curvature tensor and metric tensor on one hand, with matter represented by the energy-momentum tensor on the other. General relativity is distinguished from other theories of gravity in its coupling between matter and curvature. As is seen Einstein's original field equations did not have the cosmological constant, Λ . He introduced the cosmological term when he applied general relativity for the first time to cosmology.

⁸ As has been commonly stated: 'Space-time tells matter how to move. Matter tells space-time how to curve,' Originally quoted from Gravitation, Misner, Thorne, and Wheeler, 1973

It should be kept in mind that in the earlier part of the 20th century, a physicist's conception of the universe was much different than today. The notion of an eternal universe had been believed by humans since antiquity and there seemed no reason to doubt such a conception, even by great thinkers like Einstein. Einstein assumed that the Universe was not only static, but also spatially finite and closed. A universe with these features would remove the need to specify any boundary conditions at infinity, a feature that he found attractive. For Einstein, boundary conditions “presuppose a definite choice of the system of reference, which is contrary to the spirit of relativity” (1917a, pg. 183).

Another point he considered and took seriously was the impact gravity could have on the state of the universe. By the 20th century the notion of gravitational collapse was quite well understood, and so an emergent problem came about as to why our universe, and all the matter in it, has not gravitationally collapsed upon itself. For Einstein, it seemed as though there must have been some other force or energy distributed in the universe to provide enough outward pressure against the inward pull of gravity to keep the universe in a state of equilibrium. As it stood, Einstein's original field equations (equation 1) had the result that the universe could not remain static – a result that he was initially unwilling to accept. In response to this, Einstein inserted the cosmological constant, Λ , a quantity of energy density that was to function as an anti-gravitational force. Thus, his modified equations read as:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu} = (8\pi G/c^4)T_{\mu\nu} \quad (2)$$

Despite its ad hoc status, for a short time Λ served the purpose of keeping Einstein's universe in a steady and static state. In the paper where this was presented, “*Cosmological Considerations on the General theory of Relativity*”, Einstein states “In

order to arrive at this consistent view, we admittedly had to introduce an extension of the field equations of gravitation, which is not justified by our actual knowledge of gravitation. It has to be emphasized, however, that a positive curvature of space is given by our results, even if the supplementary term is not introduced. That term is necessary only for the purpose of making possible a quasi-static distribution of matter,” (1917a). Yet even Einstein was a little bothered by the dubious status of Λ . He once said: “I have again perpetrated something relating to the theory of gravitation that might endanger me of being committed to a madhouse.” (Letter to P. Ehrenfest, February 4th, 1917). Nevertheless, introducing Λ led to a cosmological solution that could be derived from a relatively new yet consistent theory of gravity. In spite of its dubious status, this bold step of introducing Λ could be regarded as one of the beginning steps of modern cosmology. However, in light of alternative cosmological models and soon to come evidence contrary to Einstein’s motivating conception of a stable and static cosmos, he would shortly retract Λ and come to believe that introducing it was the greatest blunder of his life⁹. The events that led Einstein to eventually denounce Λ would lead other physicists to the view that once the cat was out of the bag it would not be so easy to put it back again; this would lead to several decades of debate regarding the necessity and rightful place of this lingering and ambiguous cosmological parameter. But let us start at the beginning.

⁹ This was actually stated by Gamow who does not claim to be quoting Einstein directly, (1958, pg. 66–67). He says: “Einstein remarked to me many years ago that the cosmic repulsion idea was the biggest blunder he had made in his entire life.” The account given in Gamow’s autobiography *My World Line* (1970, pg. 44) is similar. In this respect, one cannot say for sure whether Einstein himself used the word “blunder” or whether this was Gamow’s paraphrasing.

Early Days of Λ

Prior to publishing his paper (1917a), Einstein presented a series of lectures at the Berlin Academy of Sciences where he first introduced GR publicly. He concluded his final lecture with his introduction of the equation that replaced Newton's law of gravity, (equation 2). Even though history remembers Einstein as the most famous in this respect, he was not the first to take issue with the well-established Newtonian conception of gravity. Problems with the Newtonian universe had begun to emerge toward the end of the previous century. In the 1890s both von Seeliger and Neumann noted various issues regarding standard Newtonian gravity in that it did not provide for a consistent model of a universe with an infinite space filled with a static uniform distribution of matter (see von Seeliger (1895, 1896) and Neumann (1896)). The primary issue was that in a Newtonian model of the cosmos the integrals corresponding to the gravitational potential at a point and the gravitational force at a point were divergent. Two options were contemplated in response to this issue: i) Maintain a standard theory of Newtonian gravity and conclude that a static homogeneous cosmos is simply impossible. What is possible instead would be an 'island universe' in which there is a kind of center where the mass density ρ reaches a maximum and then falls off more rapidly than $1/r^2$ as $r \rightarrow \infty$ (where r measures the distance from the center). Such a model would allow the relevant integrals to converge. And ii) Neumann proposed that the Newtonian gravitational potential $\varphi(r) = -Gm/r$ for a point mass m be multiplied by the factor $e^{(-\sqrt{\Lambda} r)}$, where Λ is a positive constant, while von Seeliger proposed to add this constant to Newton's inverse square law. This would provide a static and homogeneous solution to Newton's theory. By making Λ sufficiently small, deviations from standard Newtonian predictions

would be indistinguishable except at large distances. Thus, von Seeliger's modification implies a corresponding modification of Poisson's equation. From a small addition we go from

$$\nabla^2 \varphi = 4\pi G\rho \quad (3)$$

to

$$\nabla^2 \varphi - \Lambda \varphi = 4\pi G\rho \quad (4)$$

Even though Einstein does not refer to this maneuver in his introductory paper (1917a), it seems clear that he knew about their work since he does mention it in a previous paper (1917b) written shortly before he presented his lectures to the Berlin Academy.

The problem recognized by von Seeliger and Neumann was a good way for Einstein to introduce the cosmological term. He rejected the first option of an island universe given statistical mechanical considerations. If the island is treated as a gas whose molecules are stars, the system would diffuse and depopulate. Going along with the second option then, Einstein noted that by replacing equation (3) with equation (4), we can get:

$$\varphi = -4\pi G\rho_0/\Lambda. \quad (5)$$

Equation (5) allows for a solution that is a static uniform distribution of matter throughout space with a constant mean density of ρ_0 . It was this kind of reasoning that enabled Einstein to argue (by analogy) that his field equations could be generalized from equation (1) to equation (2). Then by assuming a homogeneous and isotropic spacetime, we obtain the most general metric:

$$ds^2 = a^2(t)d\sigma^2 - dt^2 \quad (6)$$

where $a(t)$ is the scale factor or the ‘radius of the universe’ and $d\sigma^2$ is the line element of a Riemann space of constant curvature, k : $k = 0$ (flat space), $k = +1$ (closed space), or $k = -1$ (open space). To reach a static solution we assume the stress-energy tensor must take the form of a perfect fluid where its pressure is zero¹⁰. Einstein’s original field equations (equation 1) imply that

$$d^2a/dt^2 = -(4\pi G)/3 \cdot (\rho + 3p)a \quad (7)$$

and

$$da^2 = (8\pi G\rho a^2)/3 - k \quad (8)$$

Equations (7) and (8) are the *Friedmann equations*, derived by Friedman in 1922. Of course these equations were not actually formulated until some years after Einstein’s paper in 1917, but using them here makes things easier in describing the progression of Λ . The Friedmann equations describe a dynamic universe with a changing scale factor of d^2a and da^2 , respectively. Assuming conservation of energy, equation (7) implies that a static solution would require ($d^2a = da = 0$) and that ρ is constant and positive and ($d\rho = 0$). Thus we find that

$$d\rho = -3(\rho + p) da/a. \quad (9)$$

Yet equation (7) also implies that $\rho = -3p$ and if $\rho > 0$, and so the pressure p must be negative. This would have been unpleasant for Einstein in 1917 since at that time, he was assuming a dust model, where $p = 0$. This meant that a static solution would require that ρ not only be constant but that ρ be constantly zero! If so, then the only static dust model

¹⁰ Represented formally as $T^{\mu\nu} = (\rho + p)V^\mu V^\nu + pg^{\mu\nu}$ where p is the pressure of the fluid and V^μ is the normed four-velocity of the fluid.

allowed by equation (1) is the trivially empty one. And by (8) this trivially empty model obviously cannot be spatially closed ($k = +1$).

However, to get around this problem (7) and (8) can be replaced, respectively, by

$$d^2a/dt^2 = -(4\pi G)/3 \cdot (\rho + 3p)a + (\Lambda a)/3 \quad (10)$$

and

$$da^2 = (8\pi G \rho a^2)/3 - k + (\Lambda a^2)/3 \quad (11)$$

As before, a static solution ($a = a_{Einstein} = \text{constant}$) requires a constant density. And for dust matter, equation (10) requires

$$\Lambda = 4\pi G\rho \quad (12)$$

Putting equations (11) and (12) together we get

$$\Lambda = k/a_{Einstein}^2 \quad (13)$$

By assuming $\rho > 0$, Λ will be positive as well. And so by equation (13), $k > 0$. This allows for a static universe that is spatially closed.

By introducing the cosmological term into his gravitational field equations Einstein was able to achieve a static cosmological model that, in 1917, he believed was needed to represent the actual cosmos. Even though his efforts to achieve such a model were based more on a prejudicial belief in eternal stability rather than on evidence, introducing Λ was a notably clever move. With a simple modification of his original field equations, he arrived at the static cosmological solution he was looking for. Unfortunately whatever satisfaction Einstein enjoyed from this apparent triumph would be ultimately fleeting.

Developments Against Λ

It did not take long for Λ to be met with controversy once Einstein introduced it. Various developments against it emerged in the following decade that eventually led Einstein to reject it entirely. The first issue regarding Λ came within the same year it was introduced. De Sitter (1917) produced an empty space ($T^{\mu\nu} = 0$) solution to Einstein's field equations (equation (2)) with $\Lambda > 0$. The coordinate system that De Sitter used to express the line element gave a static solution:

$$ds^2 = dr^2 + R^2 \sin^2(r/R)[d\psi^2 + \sin^2(\psi)d\theta^2] - \cos^2(r/R)dt^2 \quad (14)$$

where R is a constant (not to be confused with the Ricci curvature scalar). The De Sitter solution seemed momentarily plausible. However, it conflicted with Einstein's conception of Mach's principle¹¹ insofar as the spacetime of the model is curved without any matter-energy to account for the curvature. Maintaining the validity of Mach's principle, Einstein discarded De Sitter's solution on the grounds that it was singular. Writing to De Sitter in 1917, Einstein stated: "Such a singularity in a finite world [obviously] is, in my opinion, to be discounted as physically beyond consideration."^{12,13} But it was not long before Einstein's grounds for rejecting De Sitter's solution were weakened. The crux of Einstein's original defense collapsed when Felix Klein demonstrated that the singularity Einstein was referring to was only a coordinate artifact.

¹¹ Mach's principle here suggests that the inertia of any system is the result of the interaction of that system and the rest of the universe. This is to say, every particle in the universe has an effect on every other particle, and this principle simply would not hold in an empty universe.

¹² At the 'equator', $r = R\pi/2$, where the g_{44} component of De Sitter's metric would be zero, Einstein suspected there would be a singularity, that would essentially hide the matter necessary for Mach's principle to hold. It should be noted that the De Sitter universe was not 'wrong' per se, but simply did not describe a universe that resembled our own, and therefore was of no real cosmological significance to Einstein.

¹³ Schulmann et al. (1998, Doc. 363).

Klein's technique for showing the artificial character of the perceived singularity in the De Sitter solution was to show how to smoothly extend the De Sitter metric through the apparent singularity, but what this extension revealed however was that De Sitter's solution is not static. The originally static *appearance* was due to the fact that the De Sitter coordinates used in equation (14) are *local*, and only cover a *part* of the fully extended manifold. In essence, Klein's technique only indicated that there were no *local* singularities and this was enough for Einstein to maintain his charge against De Sitter's 'singular' solution as physically meaningless.

Around a decade later, George Lemaitre began to consider some very different cosmological implications of Einstein's field equations. He presented the new idea of an *expanding* Universe, claiming that the universe began with the explosion of a "primeval atom" (Lemaitre, 1927). He described his theory as 'the *Cosmic Egg* exploding at the moment of the creation', which was later to be coined by his critics as the 'Big Bang' theory. Lemaitre had considered cosmic rays to be the remnants of the event (although it is now known that they often originate within our local galaxy). He estimated the age of the universe to be between 10 and 20 billion years old, which is consistent with modern observations and opinion.

Lemaitre had based his *dynamic* theory of the cosmos on the work of Einstein among others. When it was made known to Einstein, he approved of Lemaitre's calculations, but refused to accept it. Lemaitre's proposal of a dynamic universe beginning from an initial singularity, once again caused a sharp reaction from Einstein similar to the one he had for Klein and De Sitter. On top of this, for others as well as Einstein, they found Lemaitre's proposal suspect because it was too strongly reminiscent

of the Christian dogma of creation and this seemed physically unjustifiable. Einstein would soon reconsider Lemaitre's theory but for a slightly different reason.

A few years later both Eddington (1930) and Lemaitre (1931) noted that Einstein's static solution was unstable. The instability arises as follows:

By setting $p = 0$ in equation (10) we end up with

$$3d^2a/dt^2 = a(\Lambda - 4\pi G\rho) \quad (15)$$

This implies that a static dust solution requires the value of Λ to be such that it exactly cancels the contribution from matter. Hence, the slightest perturbation that would make $\rho < \Lambda/4\pi G$ by any amount whatsoever would make the universe start expanding, and by equation (9) the expansion decreases the density thereby increasing d^2a , causing the universe to accelerate; an outcome that sounded much less plausible than it does today. Likewise, the slightest perturbation that leads to $\rho > \Lambda/4\pi G$ by any amount whatsoever would start a contraction leading to a cosmological crunch. At the time it was undesirable to have solutions that described an unstable universe – where 'unstable' in this case meant dynamic rather than static.

It's not known whether Einstein read Eddington's claim (1930), but he had read Friedmann's paper (1922), where Friedman showed that his field equations admit non-static solutions with a homogeneous and isotropic matter distribution (with or *without* Λ). Initially Einstein thought that Friedmann had erred in his calculations (see Einstein, 1922), but later concluded that Friedmann was correct. And in 1931, Einstein noted that the instability of his static solution did indeed follow from the Friedmann equations.

As for evidence to support these cosmological solutions, some was beginning to amount in favor of an expanding universe by the early 1920s. In 1924 Edwin Hubble

announced the discoveries he had made with the 100-inch Hooker Telescope at the Mount Wilson Observatory in California, then the world's most powerful telescope. He had noted that the fuzzy nebulae seen earlier with less powerful telescopes were not part of our galaxy as had been thought by Eddington and others, but were galaxies themselves, outside the Milky Way. Subsequently, by using redshift measurements, Hubble discovered the velocity-distance relation between the distance of galaxies from Earth and the rate of their recession¹⁴.

Hubble, like Eddington, originally interpreted his famous results on the redshift of the radiation emitted by distant 'nebulae' in the framework of the De Sitter model. However, Lemaitre's successful explanation of Hubble's discovery in terms of his expanding universe finally changed his view and Hubble published these results (1929). This is a positive instance of a 'happy family' meeting of theory and observation, similar to the afore mentioned meeting of Dicke's hypothesis of residual radio wave radiation from the Big Bang and Penzias and Wilson's discovery of CMBR. Even though Lemaitre obtained empirical support for his theory in an 'accidental' way, given that it was not specifically Hubble's intention when he made his observations, it nevertheless indicates how a cosmological hypothesis could be effectively tested.

¹⁴ Hubble's law can be stated as follows: $v = H_0 D$, where v is the receding velocity of a galaxy due to the expansion of the universe (typically measured in km/s), H_0 is Hubble's constant, and D is the current distance to the galaxy (measured in Mpc). One can derive Hubble's law mathematically if one assumes that the universe expands (or shrinks) and that the universe is *homogeneous*. For most of the second half of the 20th century the value of H_0 was estimated to be between 50 and 90 km/s/Mpc. The Hubble Key Project significantly improved the determination of the value and in May 2001 published its final estimate of 72 ± 8 km/s/Mpc. In 2003 the satellite WMAP further improved that determination to 71 ± 4 , using a completely independent method, based in the measurement of anisotropies in the cosmic microwave background radiation. (As recorded by Wikipedia: http://en.wikipedia.org/wiki/Hubble%27s_law)

Hubble's evidence helped justify Lemaitre's theory, while at the same time delivering a potential death blow to Einstein's static model of the cosmos and the necessity for Λ . It should be stated that initially, the cosmological and astronomical communities were not in complete agreement as to what Hubble's observations implied about the cosmos, but it was enough for Einstein to reconsider his original suspicion about Lemaitre's theory.

In January 1933, Lemaitre, Einstein and Hubble met in California to discuss this issue. After Lemaitre explained how Hubble's observations could accord with Einstein's theory, Einstein said, "This is the most beautiful and satisfactory explanation of creation to which I have ever listened". The work of Friedmann (1922) and Lemaitre (1927) enabled Einstein to see that Hubble's observations could be accounted for within GR in an "unforced manner," that is without the cosmological constant term (1931, pg. 237). Λ was losing its place in Einstein's field equations and in a letter to Weyl, Einstein wrote: "If there is no quasi-static world, then away with the cosmological term."¹⁵

As has been mentioned, Einstein was uneasy about introducing Λ from the beginning and once Hubble's observations had seemingly indicated that the universe is expanding he thought that " Λ was tainted with the original sin of a faulty motivation," (Earman, 2001, pg. 197). In a letter to Lemaitre Einstein further expressed these feelings as follows, "Since I have introduced this Λ term, I had always a bad conscience. But at that time I could see no other possibility to deal with the fact of the existence of a finite

¹⁵ Einstein actually wrote this letter to Weyl in 1923. Shortly after his concession to Friedmann that his calculations were indeed correct. It seems as though Einstein was prepared to remove Λ years before his 1933 meeting with Hubble and Lemaitre.

mean density of matter. I found it very ugly indeed that the field law of gravitation should be composed of two logically independent terms, which are connected by addition. About the justification of such feelings concerning logical simplicity it is difficult to argue. I cannot help to feel it strongly and I am unable to believe that such an ugly thing should be realized in nature.”¹⁶

Lemaitre makes a related point in that even if the introduction of the cosmological constant has lost its original justification as a parameter leading to a natural solution to a static universe, “it remains true that Einstein has shown that the structure of his [gravitational field] equations quite naturally allows for the presence of a second constant besides the gravitational one. This raises a problem and opens possibilities, which deserve careful discussion. The history of science provides many instances of discoveries, which have been made for reasons, which are no longer considered satisfactory. It may be that the discovery of the cosmological constant is such a case,”¹⁷ (Lemaitre, 1949, pg. 443).

Yet Einstein’s denunciation of Λ did not make it disappear entirely from cosmology. Even though its role as a ‘cosmic stabilizer’ had been debunked, some supporters opted to keep it around for its ability to help solve other cosmological problems. These efforts to keep Λ in the game sewed the seeds to what would eventually emerge as the cosmological constant problem. This point notwithstanding, some of the conservative members of the cosmological community had good reason to keep Λ alive.

¹⁶ Einstein to Lemaitre 26 September 1947 (Einstein Archives Doc. 15 085).

¹⁷ As quoted from Lemaitre’s contribution to the Schilpp volume in Einstein’s honor, (1949, 443).

Incentives for maintaining Λ

Originally, many thought that the observations of galaxy recession removed the need for Λ . Λ was only necessary to ensure a fixed and static universe, and since galaxy recession implies a dynamic universe, Λ could then be discarded. “After the discovery of the galaxy recession, the necessity of the cosmological constant became doubtful ...until recently the cosmological constant was deemed unphysical,” (Brax, 2004, pg. 228). But this is not completely true; some supporters persisted. Eddington and Lemaitre were two of the most outspoken champions of Λ , but the motivations of these two men were different as we’ll see. Despite the pre-existing controversy, there were reasons of a more positive sort for keeping Λ alive. Two main reasons were that it could help resolve the emerging issues of 1) The Age Problem of the Universe and 2) Galaxy Formation and Recession.

Regarding the Age Problem, if we take $\rho > 0$ and $p \geq 0$, it follows that for Friedmann cosmological models with $\Lambda = 0$, $d^2a > 0$ (see Equations 7 and 10). Redshift measurements tell us that $da_o > 0$ (where the naught subscript on a quantity indicates its present value). And so the present age t_o of the universe is less than the Hubble time, which is represented as:

$$\tau_{H_o} = H_o^{-1} \quad (16)$$

where

$$H_o = da/a \quad (17)$$

Hubble’s early measurements of H_o gave a value somewhere between 460 and 500 $\text{km s}^{-1}\text{Mpc}^{-1}$ ($\text{Mpc} \equiv \text{Megaparsec} = 3.3 \text{ million light years}$) which amounts to a value for τ_{H_o} of roughly two billion years. This period is not only too short compared to

the age of stars but it implies that the age of the entire universe is even younger than the Earth! Given the apparent paradox that the stars seem to be much older than the Universe, a spin-off problem came about in that our estimate of stellar ages cannot be abandoned without leaving the facts upon which it was based unexplained. This clash of ages was conveyed as a ‘problem’ by the incumbent Astronomer Royal at the time, Spencer Jones, who said “if the Universe has been in a state of progressive expansion following upon an initial disturbance when it was much more compact than at present, the longer time scale is impossible. How the two different time scales are to be reconciled is one of the outstanding problems of astronomy at present,” (Jones, General Astronomy, 1934, pg. 414).

An approach to resolving the age problem was sought by appealing to Λ . If we take Friedmann’s dust universe where $p = 0$, and the conservation laws implies that $\rho a^3 = \text{const}$, then its expansion rate can be represented with equation (11), $da^2 = (8\pi G\rho a^2)/3 - k + \Lambda a^2/3$. By setting $8\pi G\rho = C$ (a constant) and assuming a spatially closed universe ($k = +1$) we get:

$$da^2 = \Lambda a^2/3 + C/a - 1 \quad (18)$$

From here, Eddington and Lemaitre created a model from Einstein’s static universe and the big bang model where the universe expands forever in the future direction of time but it approaches the Einstein static model asymptotically in the past direction (as $t \rightarrow -\infty$). The appeal of this model is that it allows for an expanding universe to accord with Hubble’s observation while extending infinitely into the past, which would put to rest any conflicting claims between the age of the stars and the age of the cosmos.

However, as observations began to accumulate, big bang models were becoming more accepted, and a universe that became static as you go infinitely back in time seemed less realistic. In light of this, Lemaitre modified his models accordingly by setting

$$\Lambda = \Lambda_{Einstein}(1 + \delta), \quad \delta > 0 \quad (19)$$

These are big bang models that begin from a singularity and expand forever but which contain a ‘coasting’ stage where the expansion rate decreases to a minimum value when

$$a_{min} = (3C/2\Lambda)^{1/3}. \quad (20)$$

Not only does the expansion rate slow down, but if we set δ to be small enough, $a(t)$ can be made to linger indefinitely. Assuming that we are living during a time which post-dates this “coasting phase,” where $da(t) \approx 0$, the longer the coasting phase, the more the age of the universe exceeds the Hubble time. This solves the age problem without getting rid of the initial singularity. See Figure 1 below.

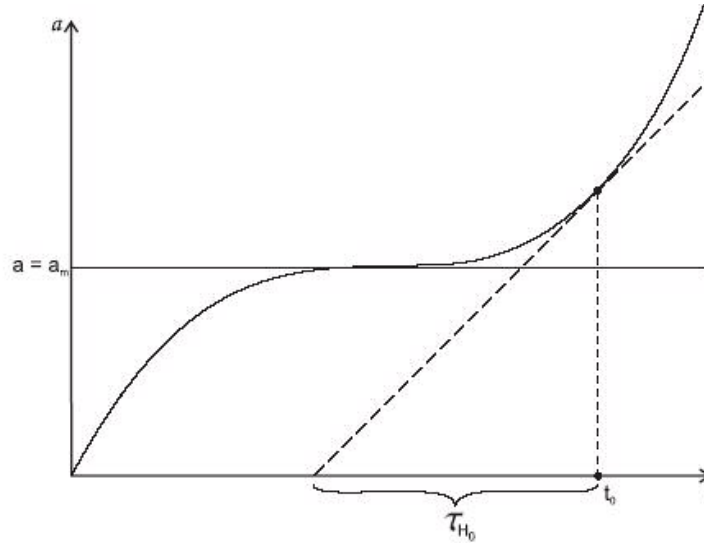


Figure 1: The so-called Lemaitre ‘hesitation’ universe can be thought of as beginning with a big bang and then proceeds with two stages of expansion. In the first stage the Λ -force is not important, the expansion rate decreases due to gravity and slowly approaches the radius of the Einstein universe. At about the same time, the repulsion exceeds the gravitational effects and a second stage of expansion begins that eventually causes the universe to expand again, the effects of which we observe. In this way a positive Λ is employed

to reconcile the age of stars with an expanding universe that emerged from a singularity, (as appears, in Earman, 2001).

In retrospect it seems like a strange set of responses to a problem in that cosmologists were willing to thoroughly manipulate their models in order to solve the age problem but seemingly never thought to approach it from a different perspective and reevaluate the basis of the problem, namely Hubble's value for H_o . Let us note that in the context within which these models were being constructed, Hubble was the only one who had a telescope capable of fixing H_o . Over the years estimates of H_o have decreased significantly, from ~ 500 to $\sim 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and the corresponding rise in the Hubble time has enabled cosmologists to recalculate the age of the cosmos and obtain a much more agreeable estimate. Because of revised values of the Hubble parameter and the development of the modern theory of stellar evolution in the 1950s, the controversy over ages was resolved. But the age problem wasn't the only incentive for maintaining a positive Λ ; there were also the problems of explaining what is causing the expansion and galaxy recession, and how large-scale structure formation occurred in the first place.

Lemaitre was an adherent of Λ not only because it offered a way around of the age problem but also because he thought it could help explain galaxy formation. Lemaitre assumed that the coasting phase in his cosmological models was preceded by density perturbations that left matter arranged in a kind of super-gas that would eventually form into dust-clouds of matter. During the coasting period, the near balance of the gravitational attraction and the repulsive force due to Λ amplified these density perturbations, causing the matter clouds to accumulate and then eventually develop into proto-galaxies and galactic clusters. At the end of the coasting phase these galactic

clusters remained condensed while the universe proceeded with its expansion, producing the recession velocities detected by Hubble, (Earman, 2001, pg. 202).

Lemaitre's intention to explain galaxy formation was notable, but as a general critique, appealing to a positive Λ as a source of density perturbations doesn't guarantee that such super structures will form in the way that we've observed them. His 'pre-coastal' density fluctuations could just as easily have been too weak to create any matter clustering or even so strong as to force the universe to collapse upon itself.¹⁸ As it stood, this was not judged to be a hypothesis persuasive enough to instigate further testing.

In a somewhat different manner, Eddington had his own reasons for thinking that Λ was essential to cosmological models. For Eddington, Λ was required to explain more than just the age problem and galaxy formation; Λ was also necessary to explain the expansion of the universe and in so doing, explain galaxy recession. In The Expanding Universe Eddington states, "I would as soon think of reverting to Newtonian theory as dropping the cosmological constant" (1933, pg. 35). For Eddington, it seemed difficult to explain how such massive structures are receding from us unless by some kind of repulsive energy source, such as a positive Λ . A model of an expanding universe where $\Lambda = 0$ would have to presuppose that large velocities of recession are built into the model from the beginning. Even if this were true, Eddington claimed, "it can scarcely be called an *explanation* of the large velocities," (1933, pg. 37). Eddington's point would later be an impetus for inflationary cosmology that attempts to minimize special initial conditions in favor of an effective cosmological constant in the very early universe.¹⁹ At the time, Eddington's concern for retaining Λ as a means of explaining galaxy recession and

¹⁸ A more thorough discussion of this critique of Lemaitre can be seen in Brecher and Silk (1969).

¹⁹ Further details on Λ 's role in inflation will be discussed in better detail in section 6.

cosmic expansion was not widely shared. It would be nearly half a century before other cosmologists would come to share his concern.

By the 1960s Lemaitre's models had lost their appeal. Due to unspecified features of the coasting phase density perturbations, whose source he did not attempt to explain, the models were left lacking causal efficacy and it was generally agreed that they did not sufficiently explain structure formation. Also given the decrease in the estimates of the Hubble constant, the age problem of the universe was put to rest. Hence, there was now less reason to appeal to Lemaitre's models, and a greater incentive to keep Λ on the shelf.

However, in late 1960s and early 1970s, Λ and these models experienced another revival. In 1967 a Λ -based Lemaitre model was temporarily revived in order to explain why quasars appeared to have redshifts that concentrated near the value $Z \approx 2$. The initial hypothesis was that quasars were born in the 'hesitation era' or coasting phase of the universe, (See Petrosian, Salpeter, and Szekeres, 1967). That way quasars at greatly different distances can have almost the same redshift, because the universe was almost static during that period.²⁰ Petrosian claimed that although a range of Lemaitre models had been ruled out by certain observations, the quasar evidence as it then existed did not rule out all such models, nor did it force a zero value for Λ . However, with more observational data, the role of the coasting phase seemed barely sufficient and wholly unnecessary which prompted Petrosian and other cosmologists to dispose of Λ for the third time. Petrosian (1974) reportedly observed several quasars with a redshift >2 . He took these to be strong evidence against the Lemaitre models, in that there was now good

²⁰ When the redshifts of emission lines in quasar spectra are 1.95, then redshifts of absorption lines in the same spectra were, as a rule, equal to 1.95. This was quite understandable, because quasar light would most likely have crossed intervening galaxies during the epoch of suspended expansion, which would result in almost identical redshifts of the absorption lines.

reason to believe there are plenty of quasars out there where $Z > 2$, and so the abundance in the $Z = 2$ range need not be anything extraordinary.

Petrosian's work is an instance when observations were done to test a Λ -hypothesis. This was a fortunate instance of experimental cosmology where the universe provided sufficient data to rule out the Λ -related hypothesis. As a result it was removed from the cosmological picture once again. Petrosian eventually concluded, "in the absence of strong evidence in favor of Lemaitre models we must once again send back the Lemaitre models and ... the cosmological constant to the shelf until their next re-appearance," (Petrosian, 1974). But particle physicists seemed less keen to comply.

These particle physicists may have been willing to rule out Lemaitre models on the basis of experiments but wanted to keep Λ around for its potential explanatory power. A new movement had emerged interpreting the cosmological constant as the vacuum energy density of quantum fields. In this respect, particle physicists were now looking for their own way to explain the observations of quasars.

The first published works that incorporate Λ into a particle physics framework were by Zel'dovich (1967, 1968). Zel'dovich had read the proposals to resurrect Λ in order to explain the $Z = 2$ redshifts of quasars (Petrosian, Salpeter, and Szekeres, 1967), and although somewhat doubtful of them, he nevertheless found them stimulating because they implied a possible connection between cosmology and particle physics; Lemaitre's idea of density perturbations in the early universe (prior to the coasting phase) could be thought of as zero-point fluctuations in a quantum vacuum. One impetus for identifying the cosmological constant with vacuum energy is because the amount of vacuum energy in the universe is also constant. Of course, one might see a problem with

this in an expanding universe: wouldn't the energy density surely decline as a given volume expands? After all, this is certainly what happens to ordinary matter. What prevents this is the peculiar equation of state of the vacuum - the work done by the pressure is just sufficient to maintain the energy density as constant. To illustrate the conservation of energy in a vacuum, consider a gas container with the piston pushed all the way in. If the energy density is ρ_{vac} , then the energy created by withdrawing the piston by a volume dV is $\rho_{vac} c^2 dV$. This must be supplied by work done by the vacuum pressure $p_{vac} dV$, and so $p_{vac} = -\rho_{vac} c^2$, or if we take $\rho > 0$, then $-p_{vac} = \rho_{vac} c^2$, making the pressure distinctly negative and therefore capable of counteracting the effects of gravity. In effect, the vacuum acts as a reservoir of unlimited energy that can supply as much as required to inflate a given region to any required size at constant energy density, (see Peacock, 1999, pg. 26). Hence, this kind of energy seemed like a good candidate for what Λ was tasked to do.

Presupposing such a connection, if Λ corresponds to the ground state energy density of the quantum vacuum, should it not be possible to calculate its value from quantum field theoretic principles? This is the turn in the history of the cosmological constant where it develops some quantum roots. By turning to QFT, some physicists were now considering the physical meaning of Λ as something different and more fundamental than how it was originally interpreted by Einstein. At this point we will switch modes and look at the quantum history of Λ . It was this alternative, quantum interpretation of Λ as vacuum energy density that, when compared with its previous cosmological interpretations, would give rise to the cosmological constant problem.

Section 3.2: The Quantum History of Λ

Early Quantum History

While the impetus to connect Λ with quantum vacuum energy density did not emerge until the 1960s, the idea that the vacuum state of a quantum system could be a potential source of energy dates back roughly half a century earlier. Inspired by the new ideas of quantum theory and Planck's law for the radiation from a black body, Walther Nernst had proposed that the vacuum is not 'empty' but is a medium filled with a large amount of energy, (see Nernst, 1916). Nernst further noted that the total energy density becomes quite large if not infinite.²¹ In 1916, this claim was neither problematic nor attractive to anyone interested in cosmology. In fact, since his interests were primarily in chemistry, Nernst's ideas about the energy content of the vacuum were largely ignored by cosmologists²².

A more solid foundation for speculations on the energy density of the quantum vacuum (ρ_{gs}) became available with the early developments in quantum field theory (see Schweber, 1994). The fundamentals of quantum field theory were developed between the late 1930s and the 1950s notably by Dirac, Fock, Pauli, Schwinger, Feynman and Dyson. Quantum Electrodynamics (QED) was the first quantum field theory in which the

²¹ Given the difficulties of handling infinite quantities, one remedy is to assume a fundamental cut-off frequency ν_0 . Even if we just consider the radiation in the vacuum to vibrate with frequencies up to, say, $\nu_0 \sim 10^{20} \text{ s}^{-1}$, the total energy content in this radiation per cubic centimeter will still be larger than $\sim 1.5 \times 10^{23}$ ergs.

²² For a more elaborate discussion of the vacuum concept before the advent of quantum ideas see the review by Saunders and Brown, 1991.

difficulties of building a consistent, fully quantum description of fields with the creation and annihilation of quantum particles were satisfactorily resolved.²³

QFT provides a theoretical framework widely used in particle physics. In particular, QED, is one of the most successful theories in physics in that it is one of the most well-tested by experiment. In a classical field theory, like classical electromagnetism (EM), there are infinitely many degrees of freedom; the electric and magnetic fields have values of $E(x,t)$ and $B(x,t)$ at each point in spacetime. The field is quantized by imposing a set of (canonical) commutation relations on the components of the electric and magnetic fields. In the quantization procedure, the classical fields are replaced by quantum operators defined at each point in spacetime. This is done by introducing field operators that define the probability of creating or destroying a particle at a particular point in space. The function of these operators is literally to create and annihilate particles within a given quantum state. In other words, a “field”, in this context, is an entity existing at every point in space that regulates the creation and annihilation of the particles. The quantization of a scalar field in (Minkowski) spacetime can be thought of as an assemblage of harmonic oscillators. The Hamiltonian operator has the form:

$$\hat{H} = \sum_k \omega_k (\hat{a}_k^\dagger \hat{a}_k + \hat{a}_k \hat{a}_k^\dagger) \quad (21)$$

where ω_k is the frequency of wave number k and \hat{a}_k^\dagger and \hat{a}_k are the associated creation and annihilation operators, respectively (as seen in Rugh and Zinkernagel, 2002). Similar to quantum harmonic oscillators, these creation and annihilation operators add and subtract energy quanta. However, contrary to a classical harmonic oscillator that can be completely at rest and have zero energy, each quantized harmonic oscillator has a non-

²³ It should be noted that there is a long standing controversy over the status of particles, namely whether to think of them as point masses or as fields. However, this controversy is not germane to the current work.

vanishing ‘zero-point’ energy when it is in its ground state. This is because in QFT, the number of particles in an area of space is not a well-defined quantity, so like other quantum observables it is represented by a probability distribution. Given that a particle maintains a certain probability of being in any specific point in spacetime, it is difficult to characterize space as being truly ‘empty’. In fact to think of space as being truly empty and the EM field as never fluctuating would lead to a violation of Hiesenberg’s uncertainty principle²⁴. By coupling the EM field with the electron-positron fields, one can speak of the creation and annihilation of virtual electron-positron pairs in the ‘interacting’ vacuum. In the general ground state of a quantum system, $|0\rangle$, it follows that $\langle 0|E|0\rangle = 0$ and $\langle 0|B|0\rangle = 0$, whereas $\langle 0|E^2|0\rangle \neq 0$ and $\langle 0|B^2|0\rangle \neq 0$. These non-zero values of the vacuum expectation values for the squared field operators can be thought of as quantum fluctuations. In contrast to the empty classical vacuum, the vacuum in QFT is an arena of seething activity. As quantum field theories have become accepted, the notion of empty space has been replaced with that of a vacuum state, defined as the ground state or the state of lowest energy density in a collection of quantum fields, and if we integrate over these quantum fields, we end up with a vacuum state that is replete with huge amounts of energy.

If the quantum vacuum is a potential source of energy, then one could wonder about its possible gravitational and thus cosmological consequences, but quantum physicists were occupied with other problems and it is not surprising that in the early years they did not worry about this possibility. Except for Pauli, who wondered in the early 1920s whether the zero-point energy of the quantum vacuum could be

²⁴ If $\Delta E \times \Delta t \geq \hbar/2$, then ΔE for any given time interval could never be zero.

gravitationally effective. Pauli had calculated that if the gravitational effect of the zero-point energies was taken into account (applying a cut-off on the zero-point energies at the classical electron radius) the world would become so compacted that its radius “would not even reach to the moon,” (Enz and Thellung, 1960, p.842)²⁵. Yet, since this back-of-the-envelope calculation was never published, Pauli’s early worries did not have much impact on the community of quantum physicists.

In the 1940’s, the issue of what kind of an impact ρ_{gs} could have for cosmology was briefly considered. In a conference address, Bohr said, “[A]ttention may also be called to the apparent paradoxes involved in the quantum theory of fields ... which imply the existence in free space of an energy density ... which ... would be far too great to conform to the basis of general relativity theory,” (Bohr, 1948, p. 222).

Bohr tersely suggests that there may be some compensation or cancellation mechanisms at work between positive and negative zero-point field energies that could lead to a solution to the vacuum energy density ‘problem’, but he thought that it would be futile to pursue such considerations at that present time. The reasoning may have been that the potential enormity of ρ_{gs} could be a mere mathematical artifact of QFT and was not something that needed to be taken all too seriously. Weinberg (1989) has claimed that since the cosmological upper bound on Λ was vastly less than any value expected from particle theory, most particle theorists at the time most likely assumed that for some unknown reason this quantity was zero. It should also be noted that at this time there was

²⁵ Pauli’s concern with zero-point energies is also mentioned in Straumann’s paper “*The History of the Cosmological Constant Problem*,” (Straumann, 2002). In it he rederives Pauli’s result by inserting the calculated energy density of the vacuum into an equation that relates the universe’s radius of curvature with its energy density, $\rho \sim 1/a^2$. The numerical value was derived from Einstein’s equations for a static dust filled universe and when the constants are accommodated, the result is that the radius of the universe is about 31 km - much less than the distance to the moon to say the least.

very little observational evidence in favor of ρ_{gs} . However, around the time Bohr made his remarks, the first evidence for vacuum energy was being obtained, namely the results of H.B.G. Cassimir that Bohr may not have been aware of.

In 1948 Cassimir predicted that due to ρ_{gs} , two uncharged parallel metal plates will have an attractive force drawing them together. As was observed during an experiment at the Phillips Research Labs in Holland, two perfectly conducting parallel plates were attracted to one another with a force proportional to $1/d^4$, where d is the separation distance, (in a Minkowski space)²⁶. This force is only measurable when the distance between the two plates is extremely small, on the order of several atomic diameters. The usual explanation begins with the presupposition that space is filled with vacuum fluctuations, where virtual particle-antiparticle pairs continually form out of nothing and then vanish back into nothing an instant later. The gap between the two plates restricts the range of wavelengths possible for these virtual particles, and so fewer virtual modes exist within this space. This results in a lower energy density between the two plates than is present in open space; in essence, the vacuum energy density between the two plates is lower than outside, causing a force to draw the plates towards each other. The narrower the gap, the more restricted the vacuum modes and the smaller the vacuum energy density, and thus the stronger is the attractive force. But this result would not become cosmologically relevant for another two decades.

Rugh and Zinkernagel give a very thorough account of the early history ρ_{vac} in their paper “*The Quantum Vacuum and the Cosmological Constant Problem*,” (2002). In it they say that with respect to the early years, “the possible large energy content of the

²⁶ Whether this effect could still be detected in a curved spacetime such as the kind of manifold represented in GR is debatable, which raises questions of how a quantum vacuum state is to be identified in such a manifold and how QFT is to be done in curved space; a non-trivial problem.

vacuum from QFT does not seem to have played any role in the discussions of the cosmological constant in the cosmology literature,” (pg. 668). This is true in a sense, but ideas relating density fluctuations with Λ date back to the early work of Lemaitre (1934). As mentioned in the previous subsection, the Lemaitre models relied on density perturbations in the early universe (prior to the coasting phase) in order to explain the final result of galaxy formation. Even though Lemaitre did not speak of these density fluctuations as arising from a quantum vacuum state, he did say, “everything happens as though the energy *in vacuo* would be different from zero,” (1934, pg. 12). Lemaitre derived a value for the energy density of the vacuum, ρc^2 . Assuming a dust-filled universe and appealing to equation (12) we get:

$$\rho_{vac} = \Lambda c^2 / 4\pi G \approx 10^{-27} \text{ g/cm}^3. \quad (22)$$

Lemaitre’s estimate of the energy density of the vacuum is only slightly less restrictive than the constraint that accords with our current observations (by two orders of magnitude). Although Lemaitre provided a physical interpretation of Λ , it was not characterized in terms of the *quantum* vacuum that was occupying physicists at the time, i.e., Planck’s constant, \hbar , does not appear in his discussion of the vacuum energy density²⁷. This was nevertheless the beginning of a connection between the cosmological constant and vacuum energy and a significant step toward considering the possibility that dynamics on the micro level at the beginning of the universe could have profound effects on the universe’s development. But the rest of the physics community was not so quick to pick up where Lemaitre left off. For decades hardly anyone seemed to have worried

²⁷ This point notwithstanding, Lemaitre may have been aware of the vacuum energy arising in quantum field theory. In a 1933 article, ‘*The Uncertainty of the Electromagnetic Field of a Particle*,’ he discusses Heisenberg uncertainty relations for the EM field. (See Lemaitre, 1933).

about how or if quantum fluctuations contributed to the cosmological constant. It wasn't until the 1960s that some physicists thought that the quantum vacuum could be used to do some real explanatory work and this would have ramifications that extend from the subatomic realm to the furthestmost reaches of the universe.

Recent Quantum History

The first published discussion of the contribution of quantum fluctuations to the cosmological constant was Zel'dovich's 1967 paper, "*The Cosmological Constant and Elementary Particles*." As mentioned earlier, it was believed at that time that a non-zero Λ was needed to account for observations of quasars. By this time, Hubble's constant had been revised and reduced and observational constraints on the value of Λ had already been calculated and published, (see Petrosian, Salpeter, and Szekeres, 1967). Zel'dovich reasoned that if we take Hubble's constant, H_0 , to be $\sim 65 - 70 \text{ km s}^{-1} \text{Mpc}^{-1}$, and

$$H_0 \approx 1/t_0 \tag{23}$$

we obtain a value for the age of the current universe as

$$t_0 \sim 1.5 \times 10^{10} \text{ years}. \tag{24}$$

And if we take the radius of the current universe, a_0 , to be

$$a_0 = ct_0 \tag{25}$$

we obtain a value

$$a_0 \approx 10^{28} \text{ cm}. \tag{26}$$

If we then consider that

$$a_0 \sim \Lambda^{-1/2} \quad \text{or} \quad \Lambda \sim 1/a_0^2 \tag{27}$$

we obtain an observational constraint:

$$\Lambda = 10^{-56} \text{ cm}^{-2} \quad (28)$$

This was the constraint that Zel'dovich worked with. Taking this value for Λ and putting it into the same equation Lemaitre used to calculate ρ_{vac} , equation (22), we get

$$\rho_{vac} \approx 2 \times 10^{-29} \text{ g/cm}^3 \quad (29)$$

And in terms of *energy* density, ρc^2 , we get

$$\rho_{vac} \approx 2 \times 10^{-8} \text{ erg/cm}^3 \quad (30)$$

Zel'dovich (1967) began by asking how one is to visualize a theory in which such properties of the vacuum are obtained from our notions regarding elementary particles. In an attempt to *derive* a value for ρ_{vac} Zel'dovich says, “The starting point of such a theory are the formulas that give the required order of magnitude of $[\rho_0]$, expressed in terms of constants m , c , \hbar , and G , where m is the elementary particle mass [proton mass],” (1967, p. 316). He shows that by using the formulas of Eddington (1931) and Dirac (1938) for the quantities characterizing the contemporary universe, and the connection between these quantities and Λ , we obtain

$$\rho_{gs} = \rho_{vac} = Gm^6 c^4 / \hbar^4. \quad (31)$$

By introducing the Compton wavelength of a proton,

$$\lambda_p = \hbar / (m_p c) \quad (32)$$

equation (31) can be rewritten as:

$$\rho_{gs} = \rho_{vac} = Gm^2 / \lambda \times 1/\lambda^3 \quad (33)$$

This latter formula corresponds to the assumption that the vacuum contains virtual pairs of particles with effective density, $n \sim 1/\lambda^3$. Zel'dovich does not address why the zero-point energies of the fields do not build up a huge Λ ; instead he assumes that the zero-

point energies, as well as higher order EM corrections to this, are effectively cancelled to zero in the theory. But even with this assumption, if the zero-point contributions to the vacuum energy density are exactly cancelled like this, there still remain higher-order effects. In particular, he thought one must accept the possibility of gravitational interactions between the particles in the vacuum fluctuations. The energy of the gravitational interaction of these pairs (Gm^2/λ) does *not* vanish to zero. Numerically, equation (33) with mass equal to m_{proton} “yields a value 10^8 times larger than required,” (Zel’dovich, 1967, p. 317). And so the essence of Zel’dovich’s paper is that this ‘left over’ vacuum energy, acting as a Λ -term would cause the universe to expand in a way that could explain the quasar observations. For the sake of giving a quantum field theoretic basis to the observations and thus a connection between particle physics and cosmology, Zel’dovich refrained from sending Λ back to the shelf.

The next year, Zel’dovich published another paper where he emphasizes more strongly than in his previous paper that zero-point energies of particle physics theories cannot be ignored when gravitation is taken into account (1968). He obtains a value for QED zero-point energy by the following formula:

$$\rho_{vac} = m(mc/h)^3 \sim 10^{17} \text{ g/cm}^3 \quad (34)$$

and so by equation (22) we get

$$\Lambda = 10^{-10} \text{ cm}^{-2} \quad (35)$$

Where ‘ m ’ is taken as the proton mass. He further notes that since this estimate exceeds observational bounds by 46 orders of magnitude it is clear that ‘...such an estimate has nothing in common with reality,’ (p.392). Even though Zel’dovich didn’t know it, by explicitly discussing the discrepancy between estimates of quantum vacuum energy and

observational constraints, he set the stage for the cosmological constant problem. This was the beginning of a minor industry that consists of trying to estimate the value of the vacuum energy density that is consistent with observational constraints. When further calculations are done on the Planck scale, the discrepancy increases from the mere 46 orders of magnitude that Zel'dovich noted, to 120 orders of magnitude! With the bumpy history of the cosmological constant laid out thus far, let us now look at the cosmological constant problem in more detail.

Section 4: Development of the Cosmological Constant Problem

Section 4.1: The Rise of the Cosmological Constant Problem

The cosmological constant problem began with a mismatch between the typical energy scales considered by particle physicists on one hand and the observational constraints imposed by cosmologists on the other. No doubt that with the increased attention that quantum vacuum energy has received from particle physicists and their persistence in interpreting Λ as the accumulation of that vacuum energy, particle physics has made a strong impact on cosmology and the status of Λ . Brax claims that “one must realize that the cosmological constant is nothing but the vacuum energy.” (2004, pg. 231). This is the Identity view aforementioned in the introduction (more on this in the next section). It is by equating Λ with ρ_{gs} that the cosmological constant problem has arisen.

As the previous sections have conveyed, Λ has been in and out of Einstein’s equations ever since he introduced it in 1917. Just as Λ ’s physical status, meaning, and role in cosmology have been interpreted and reinterpreted over the decades, so too has the meaning of ρ_{gs} . Perhaps some further remarks about ρ_{gs} would be worth noting here since our understanding of it is critical in how we are to interpret the cosmological constant problem.

Further Developments of Vacuum Energy

It is granted that in quantum field theories that underlie modern particle physics, the notion of empty space has been replaced with that of a vacuum state in a collection of quantum fields. In QED, we conceptualize quantum fields as being infinite in extent, but

if we wish to estimate the field strength within a finite volume (even if that volume is the entire universe), we introduce a standard ‘box quantization’ procedure. This entails a ‘quantization volume’ V , where we can exploit the formal equivalence of a field mode with a harmonic oscillator. By representing the EM field as a set of harmonic oscillators, we can derive a finite expression for ρ_{vac} from the sum of the zero-point energies for each oscillator mode. So,

$$\rho_{vac} = E/V = (1/V) \sum_k \frac{1}{2} \hbar \omega_k \quad (36)$$

where ω is the oscillation frequency of the corresponding classical harmonic oscillator, and k refers to frequencies and wave-numbers of a continuum of (plane-wave) modes.

This technique was well understood by the middle of the 20th century and there was good reason at this time to think that the quantum vacuum energy was physically real since the effects of ρ_{vac} had been tested and observed – the Cassimir experiment had indicated that the virtual particles leave a detectable trace of their activities. And in the spirit of classic experimentation, this was quite convincing.²⁸ Even if we were to grant the experimental evidence as a valid indicator for the existence of quantum vacuum fluctuations, we must mindfully consider how ρ_{vac} relates to Λ . Is the energy density represented by Λ simply an accumulation of vacuum energy? We can say for the time being that these fluctuations *may* be the source of the energy and effects we attribute to Λ , but by admitting this we dive into the heart of the cosmological constant problem. As previously stated, advocates of QFT have calculated the value of ρ_{vac} to be much higher than its observationally constrained value. Before we can say whether these fluctuations

²⁸ Disputes of these claims will come later in section 5.3

are the physical basis of Λ , it's worth reviewing the other numerical estimates for ρ_{vac} that have emerged out of QFT research aside from what was derived by Zel'dovich.

Development of the Standard Model

In 1968, Zel'dovich arrived at a 'natural' value for ρ_{vac} of $\sim 10^{17} \text{ g/cm}^3$ (recall equation (34)). Since that time the 'Standard Model' of elementary particle physics has developed. Particle physicists consider the Standard Model a highly accurate model, insofar as it accurately describes microphysics up to energies of the order of $\sim 100 \text{ GeV}$. Within the Standard Model, matter is made up of leptons and quarks that interact through three basic types of interactions: the electromagnetic, the weak and the strong interactions. Strong interactions are governed by the theory of quantum chromodynamics (QCD), comprising its own sector of the Standard Model. The electromagnetic and weak interactions are governed by and unified in the electroweak theory. The Standard Model includes an additional coupling of its constituents (or fields) to Higgs fields that play a critical role both in formulating the electroweak theory, and in generating the masses of the particles. These features are relevant to cosmology in that they describe microphysical dynamics that were allegedly occurring at different stages in the early universe and from them we can use the standard model to derive energy estimates that contribute to the ground state in the early universe.

It should be noted that the Standard model has its problems too. The primary ones being that the model has several free parameters that cannot be calculated independently, and it does not describe gravitational interactions. For these reasons, the Standard Model is seen as adequate for what it describes but nevertheless incomplete. Yet if the Standard

Model is to be taken seriously, then we arrive at estimates for ρ_{vac} to be much greater than Zel'dovich's 10^{17} g/cm^3 (or $\sim 10^{28} \text{ erg/cm}^3$). So how large is ρ_{vac} ? The numerical answer depends on which frequency interval we employ in the summation in equation (36).

Physicists have extended the range of the EM field modes from zero up to an ultraviolet cut-off set by the electroweak scale $\sim 100 \text{ GeV}$, where the EM forces become effectively unified with the weak forces in the more general framework of the electroweak interaction. At this energy range, a numerical value for the zero-point energy could be roughly estimated as

$$\rho_{vac}^{EW} \approx (100 \text{ GeV})^4 \approx 10^{48} \text{ erg/cm}^3 \quad (37)$$

This is already an immense amount of vacuum energy attributed to the QED ground state which exceeds the observational bound (equations (29) and (30)) by ~ 56 orders of magnitude. In our units of $\hbar = c = 1$, a characteristic energy of $E \text{ (GeV)}$ translates into a characteristic energy density of $E^4 \text{ (GeV}^4\text{)}$, or $\rho \sim E^4$. And for concreteness, as a reminder, $1 \text{ eV} = 1.6 \times 10^{-12} \text{ ergs}$ and $1 \text{ eV} = 1.97 \times 10^{-5} \text{ cm}^{-1}$. So in natural units the energy density in equation (30) can be represented as:

$$\sim 10^{-8} \text{ erg/cm}^3 \text{ or } (10^{-12} \text{ GeV})^4, \text{ or } \sim 10^{-48} \text{ GeV}^4 \quad (30')$$

Taking into account that a particle physics model is only valid within a certain energy range, to extrapolate quantum field theories beyond the energies of the electroweak scale of $\sim 100 \text{ GeV}$ involves an element of speculation, at least at present. However, new particle accelerator technologies such as CERN's Large Hadron Collider are being developed that will hopefully enable physicists in the not too distant future to probe the kinds of interactions that occur beyond this energy range. Many physicists nevertheless

maintain the belief that the QFT framework is effectively valid up to scales set by the Planck energy²⁹, corresponding to the cosmological epoch when the universe was 10^{-43} seconds old. At the Planck scale,

$$E_{Planck} = (\hbar c^5/G)^{1/2} \approx 10^{18} \text{ GeV} \quad (38)$$

If we assume that energy on the Planck scale can originate from QED zero-point fluctuations, we get

$$\rho_{vac}^{Planck} \approx (10^{18} \text{ GeV})^4 \approx 10^{72} \text{ GeV}^4 \approx 10^{110} \text{ erg/cm}^3 \quad (39)$$

On this scale, the estimated value for ρ_{vac} has thus become inconsistent with the observational constraint of equation (30) by ~ 120 orders of magnitude. Whether or not we take QFT to be a valid theory of particle physics at the Planck scale, we are still left with a major discrepancy of many, many orders of magnitude that would require explaining. After the Standard Model was established, the original problem with the cosmological constant that Zel'dovich pointed out had now become a bigger problem. Much of the work done to resolve the problem since then has been in trying to determine some cancellation or depletion mechanism to explain how such a huge amount of energy supposedly vanished. Tackling this problem was thought easier by assuming that Λ was cancelled out to exactly zero.

²⁹ Albeit speculative, perhaps one day a new physics will emerge between the electroweak and Planck scales. If such new physics remains quantum field theoretical in nature, then we may assume that there will still be vacuum energy at this scale. In fact, QFT physics between the Planck and Electro-weak scales would be required if we want a numerical value for ρ_{vac} to be physically meaningful. It is between these scales and between these stages in the early universe where inflation occurs, which, according to most inflationary models relies on ρ_{vac} as its driving force.

Section 4.2: Interpreting the Cosmological Constant Problem

One of the appealing features of the electroweak theory was that it could effectively describe the dynamics of particles (bosons and fermions) from the vacuum state into various other excited states, but in order for this theory to describe such dynamics, one needs to introduce a Higgs field that gives masses to the particles by means of ‘spontaneous symmetry breaking’³⁰. The worries about vacuum energy took a new turn when it was realized that this spontaneous symmetry breaking mechanism invoked in the electroweak theory might have cosmological consequences.

Generally speaking, the governing equations for the dynamics of the fields have a symmetry that is not shared by the vacuum state; the vacuum state then breaks the symmetry in question. A common example to help clarify this phenomenon is a ball sitting on top of a mound. This system is in a completely symmetric state, but it is not a stable one; at some point, the ball can spontaneously roll down the hill in one direction or another. When this happens, the symmetry will have been broken because the direction the ball rolled down has now been singled out from other directions. See figure 2 below.

³⁰ All the massive particles in the Standard Model are coupled to the Higgs field and their masses are proportional to the vacuum expectation value of the Higgs field, which is non-zero in the broken phase (see Weinberg, 1995 for a historical account of spontaneous symmetry breaking, pg. 676).

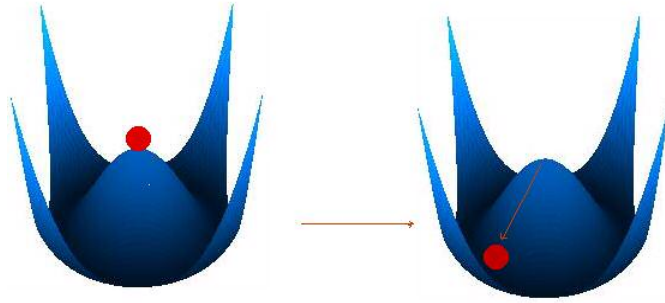


Figure 2 This diagram is a modified version of the one that appears at: http://en.wikipedia.org/wiki/Spontaneous_symmetry_breaking. In physics, one way of understanding spontaneous symmetry breaking is with a Lagrangian, which basically dictates how a system will behave. A Lagrangian can be split into its kinetic and potential terms. It is in the potential term that the action of the symmetry breaking occurs. The potential has many possible minimum states, or vacuum states. Once the system chooses a minimum state to roll into, the symmetry is broken.

This symmetry-breaking process relates to cosmology since within the Big Bang framework it is assumed that the universe expands and cools and during the cooling process the universe passes through some critical temperatures corresponding to characteristic energy scales that in turn correspond to cosmological phase transitions.³¹ These transitions imply symmetry breakings where the vacuum state of the quantum fields becomes less symmetric than before. A general picture emerges of a chain of symmetry breakings in the earlier phases of the universe producing a much less symmetric vacuum state at present. It is believed that shortly after the big bang (roughly 10^{-35} seconds after), the chain of symmetry breakings began and is commonly conceived as follows:

(GUT symmetry breaking, $E \approx 10^{14} \text{ GeV}$, $t \approx 10^{-35} \text{ s}$ after big bang) \rightarrow

(E-weak symmetry breaking, $E \approx 10^2 \text{ GeV}$, $t \approx 10^{-11} \text{ s}$ after big bang) \rightarrow

(QCD symmetry breaking, $E \approx 10^1 \text{ GeV}$, $t \approx 10^{-3} \text{ s}$ after big bang)

³¹ This happens because the Hamiltonian of a system usually exhibits all the possible symmetries of the system. Yet the low-energy states lack some of these symmetries, at low temperatures, the system tends to be confined to the low-energy states. At higher temperatures, thermal fluctuations allow the system to access states in a broader range of energy, and thus more of the symmetries of the Hamiltonian.

The first of these phase transitions has not been mentioned because it occurs at a time before (or beyond rather) the scope of the well-established Standard Model. The first symmetry breaking represents the phase transition associated with the grand unified theory (GUT). The grand unified symmetry group then breaks down into the symmetry group of the Standard Model. Note that one has very little, if any, experimental evidence to constrain theoretical estimates above energy scales of ~ 100 GeV. These phase transitions at these energy scales are still speculative. Nevertheless, the approximate energy scales where these symmetry breakings are thought to occur can then be translated into expectations for vacuum energy densities as, $\sim (10^{14} \text{ GeV})^4$, $\sim (10^2 \text{ GeV})^4$, and $\sim (10^{-1} \text{ GeV})^4$, respectively. Rugh and Zinkernagel note that “Under the assumption that vacuum energy density can be identified with a cosmological constant, this implies a hierarchy of different cosmological constants, one for each phase of the vacuum state,” (2002, p. 45).

Some of the authors who first considered the cosmological impact of spontaneous symmetry breaking (Linde (1974), Dreitlein (1975), and Veltman (1975)) did worry about vacuum energy and the cosmological constant, but they did not express such worries in terms of a cosmological constant *problem*. Linde noted that without spontaneous symmetry breaking “the ‘old’ theories of elementary particles have yielded no information whatever on the value of Λ ,” (1974). For Linde, spontaneous symmetry breaking was essential for any explanatory account of why Λ (or the present vacuum state) has the numerical value it has. What is essential is the *difference* of vacuum energy density before and after symmetry breaking. Linde took spontaneous symmetry breaking

to imply that the vacuum energy density depended on the state of the universe, namely its temperature. Hence, in a hot Big Bang universe where the temperature decreases with time, almost the entire change in ρ_{vac} occurred at the time of the symmetry breakings. He estimated that, given the observational constraints, the vacuum energy density must have changed at least 50 or more orders of magnitude from times before the electro-weak spontaneous symmetry breaking until today, (compare equations 30 and 37). Spontaneous symmetry breaking seemed like a plausible mechanism for cutting down the value of Λ , but whether it would reduce its value enough to make it consistent with observational constraints was still uncertain.

By the mid-1970s most physicists were in agreement that spontaneous symmetry breaking occurred in the early universe, but there was much disagreement regarding its impact on Λ . Some thought that if the symmetry of the universe was broken, the kind of cancellation necessary to bring ρ_{vac} down to zero and in line with observational constraints could not occur, while others thought that it was through spontaneous symmetry breaking that ρ_{vac} could be sufficiently reduced. Some thought spontaneous symmetry breaking created the cosmological constant problem, some thought it solved the problem and some thought it had nothing to do with anything.

A few years after Linde's paper (1974) was published, Bludman and Ruderman (1977) published a paper that de-emphasized the efficacy of spontaneous symmetry breaking. In it they argued that even if the vacuum energy density was very large at the time of the symmetry breaking, it was comparatively negligible to the thermal energy density of ultra-relativistic particles present at that time. They further noted that the effect of this thermal energy density was to smooth out any effects of a large vacuum energy

density and in their view there was no hope of either confirming or refuting the spontaneous symmetry breaking hypothesis by means of observational constraints on Λ . Bludman and Ruderman conclude, “To this old problem (or pseudo-problem), neither broken symmetry nor we have anything to add,” (1977, p. 256).

Bludman and Ruderman called the problem of the cosmological constant a ‘pseudo-problem’ since we could plausibly arrive at any value of Λ by adding in the right counter terms, and the inconsistency with the observed constraints on Λ is a problem only if one believes, as Zel’dovich did, that it should be derivable from other fundamental constants in particle physics. They did not elaborate on this point but they made a reference to Zel’dovich’s 1968 paper in which it is suggested that a possible relation between Λ and fundamental constants in particle physics might be “useful in the construction of a genuine logically-consistent theory,” (pg.384). For Bludman and Ruderman, Λ might not be derivable from a more fundamental theory that incorporates both gravity and particle interactions. It further suggested that spontaneous symmetry breaking did not necessarily resolve the cosmological constant issue nor did it create a consensus that the cosmological consequences of ρ_{gs} , if any, had resulted in a critical problem for modern physics.

For those who shared their view, perhaps the cosmological constant problem was merely a pseudo-problem for determined physicists looking for connections where there were none. Zel’dovich tried to derive a value for Λ using fundamental constants in particle physics, and did not arrive at any acceptable result. Straumann has similarly remarked upon the ‘unnaturalness’ of Λ , but sees the problem as more significant than Bludman and Ruderman. He says, “Trying to arrange the cosmological constant to be

zero is unnatural in a technical sense. It is like enforcing a particle to be massless, by fine-tuning the parameters of the theory when there is no symmetry principle, which implies a vanishing mass ... This problem is particularly severe in field theories with spontaneous symmetry breaking. In such models there are usually several possible vacuum states with different energy densities,” (Straumann, 2004, p.10). Straumann recognized that symmetry breaking can lead to different values for Λ and this makes it difficult to arrive at one specific value from a unique arrangement of constants.

For Bludman and Ruderman, since thermal fluctuations had such a dominating effect in the early universe, vacuum fluctuations are not something that warranted serious attention. And so if the cosmological constant as ρ_{vac} was nothing to be concerned with, then the cosmological constant *problem* was nothing to be concerned with. This position of Bludman and Ruderman did not succeed in swaying the interested parties into accepting the cosmological constant problem as a pseudo-problem; since it is still being debated today. This may be due to an extended interest in the explanatory power of vacuum energy.

The advent of inflationary cosmology in the early 1980s stimulated further interest in the cosmological effects of vacuum energy. Inflation, a short period of rapid expansion in the early universe driven by vacuum energy, was an idea that was introduced to resolve several issues that were hampering standard big bang cosmology at that time, (see Guth, 1981). The idea claimed to have provided cosmology with possible solutions to I) the horizon problem, II) the flatness problem, and III) the monopole problem. These problems concern, respectively, the question of why the background radiation is isotropic, why the observed universe is Euclidean to a very high degree, and

why no monopoles are observed (these monopoles are objects which are expected on GUT grounds).³² If cosmologists at this time wished to hold onto inflation as an explanatory mechanism, it would have been difficult to dismiss the role of Λ and the necessity of spontaneous symmetry breaking. This changed the status of the problem slightly.

Before inflation was introduced, some cosmologists were content keeping Λ on the shelf because it was not currently needed to solve any pressing cosmological puzzles. At that time, it was the particle physicists who, by identifying Λ with ρ_{vac} , sought a physical explanation of why $\Lambda \approx 0$. When inflation was proposed (Guth, 1981), it required the services of Λ once again, but this time Λ was explicitly identified with a large quantity of vacuum energy necessary to drive the exponential expansion. Λ was brought back into Einstein's field equations not as an ad hoc energy parameter as Einstein had introduced it originally, but as vacuum energy. Since inflation is thought to be driven by vacuum energy, inflationary models required a large amount of it during the GUT phase transition (at least this was the original idea). The inflationary picture fit well with the idea of spontaneous symmetry breaking, and it seemed plausible that Λ may have vanished over a series of subsequent symmetry breakings after inflation. The cosmological constant problem had become a problem of explaining the vanishing of ρ_{vac} only after it's done the work we want it to do. Originally, trying to resolve the cosmological constant problem meant trying to derive a value for the vacuum energy density that was zero or close to it. Now there was a greater emphasis on ensuring that

³² For a comprehensive analysis on inflation, see Earman and Mosterin (1999).

ρ_{vac} did not start out as zero, but ended up that way after inflation. Λ had been brought back to do some explanatory work once again.

Until this time, physicists had considered the cosmological constant problem to be either a pseudo-problem or one of an epistemic nature, where a mechanism such as spontaneous symmetry breaking or something like it was responsible for the vanishing of Λ , and we just had not worked out the details yet. Even if it took some fine-tuning, Λ ought to have been zero. As the history of the problem indicates, it seems that its status changes as the status of Λ changes, which often happens when advances are made in cosmology. In the 1990s, various observations were made that had a significant bearing on Λ 's role in cosmology, which in turn had a bearing on the cosmological constant problem.

Section 4.3: Recent Observations and Relevant Cosmological Facts

As an experimental science, cosmology is no longer a data-starved field with loose or non-existent constraints on the basic parameters that characterize cosmological models. Observations have been made and data collected that allow cosmologists to form testable hypotheses. Recent observations of type 1a Supernovae in conjunction with anisotropies in the cosmic microwave background radiation (CMBR) have led cosmologists to reconsider the contents of the cosmos and the way in which it is expanding. By reinterpreting these features of the cosmos, Λ 's role has taken on a new significance, as we shall see. Some cosmologists have come to believe that not only is $\Lambda \neq 0$ but that it is affecting the state of the current cosmos.

Type Ia supernovae were originally designated as ‘standard candles’, in that they have a distinct luminosity signature that can be identified at various distances. Their intrinsic brightness can reliably function as a standard light source by which distances can be gauged. Fortunately for us, nature has been generous enough to provide us with many of these standard candles that have allowed us to develop a solid understanding of their behavior, which in turn allow for significant progress to be made in determining various features of the cosmos. In the 1990s they were originally employed for a ‘parameter determining’ experiment, namely to get an accurate estimate of the Hubble expansion rate. This was done by finding such supernovae beyond our local cluster of galaxies. The farther away the supernovae, the earlier you can measure the expansion rate. It was originally thought that making such measurements would enable us to determine the *deceleration* of the expansion caused by gravity. Since the deceleration would depend on the cosmic mass density, to measure the expansion rate would, in effect, enable us to ‘weigh’ the universe. The greater the density, the more the expansion is slowed by gravity. As true as this would have been for the simplest of cosmological models, such models did not accord with observation. In the late 1990s the Supernova Cosmology Project at Berkley set out to make the observations necessary to determine this deceleration rate. Much to these researchers’ surprise, the data indicated that the universe was not decelerating at all, but rather it is accelerating! In this cosmological experiment, the data disconfirmed the hypothesis in question as effectively as any other kind of ‘classic’ experiment³³. The experimenters began claiming that if the universe

³³ However these initial data were subject to doubts due to extinction effects (due to the presence of cosmic dust), and possible evolutionary effects of the Type Ia supernovae, but further Supernovae experiments have been done and are being planned in the future to further support these initial findings.

were decelerating we would see a very specific distribution of supernovae. To test if this was the case, data were obtained by pointing their telescopes in various directions. Each direction could be thought of as a different instance of testing an idea under slightly varied conditions. The result was that despite the variation, the expected distribution of supernovae, or expected relation between supernovae and red shift, was simply not there. The hypothesis of the decelerating universe had been falsified.

Recent red shift vs. distance measurements on Type Ia supernovae indicate that the deceleration parameter, q_o , is indeed negative for these distant objects (see Riess et al. (1998), Schmidt et al. (1998), and Perlmutter et al. (1999)). The apparent brightness of these standard candles is much less than would be expected if $q_o \geq 0$. Hence, the simplest models were too simple. The data fit quite well with a modified model of an expanding universe that is being accelerated by some kind of energy source. The best fit implies that in the present epoch the energy density is larger than the matter density. In fact, other observations have indicated that the majority of the universe, roughly 70%, is composed of this energy. The exact nature of this energy is still a matter of speculation, so cosmologists have come to call it ‘dark’ energy. It is best described as a hypothetical form of energy that permeates all of space and has strong negative pressure that effectively acts in opposition to gravity at large scales causing the cosmic expansion rate to accelerate. The two proposed candidates for dark energy are: 1) Λ , a constant quantity of vacuum energy density filling space homogeneously, and 2) Quintessence, a *dynamic* field whose energy density can vary in time and space. How these candidates differ will be discussed in greater detail in the following sections. Suffice it to say, if these two kinds of energy are different, the schematics involved in explaining the recent

observations would be different. If we were to assume for the moment that Λ was responsible for these observations, the relevant schematics would be as follows. The cosmic deceleration factor, q , is defined in a Friedmann model as

$$q = \frac{1}{2} \Omega^M (1 + 3p/\rho) - \Omega^\Lambda \quad (40)$$

where Ω^M and Ω^Λ are the dimensionless density parameters associated with matter and Λ , defined respectively as:

$$\Omega^M = 8\pi G\rho/(3H^2) \quad , \quad \Omega^\Lambda = \Lambda/(3H^2) \quad (41)$$

When $\Lambda \leq 0$ and $(\rho > 0, p \geq 0)$, then $q > 0$ and the rate of expansion must slow down. If $q < 0$, and we're speeding up, then one must have either a big enough $\Lambda > 0$ or ... a strange form of matter. If we ignore strange forms of matter for the moment³⁴, the present matter dominated epoch suggests

$$q_o \approx \frac{1}{2}\Omega_o^M - \Omega_o^\Lambda \quad (42)$$

If we take $k = 0$, the case of a flat universe preferred by most inflationary models, then

$$\Omega_o^{Total} = \Omega_o^M + \Omega_o^\Lambda \approx 1 \quad (43)$$

Current observational evidence suggests that $\Omega_o^M \approx 0.3$, and if the universe is more or less spatially flat, as recent CMBR measurements strongly suggest, then $\Omega_o^\Lambda \approx 0.7$. From equations (42) and (43) we can derive

$$q_o \approx \frac{1}{2} - \frac{3}{2} \Omega_o^\Lambda. \quad (44)$$

Inserting the value of Ω_o^Λ into equation (44), we get

$$q_o \approx -0.55 \quad (45)$$

so that the expansion rate of the universe is indeed accelerating. See figure 3 below.

³⁴ We will address this theme in section 6.1, when we look at Quintessence models.

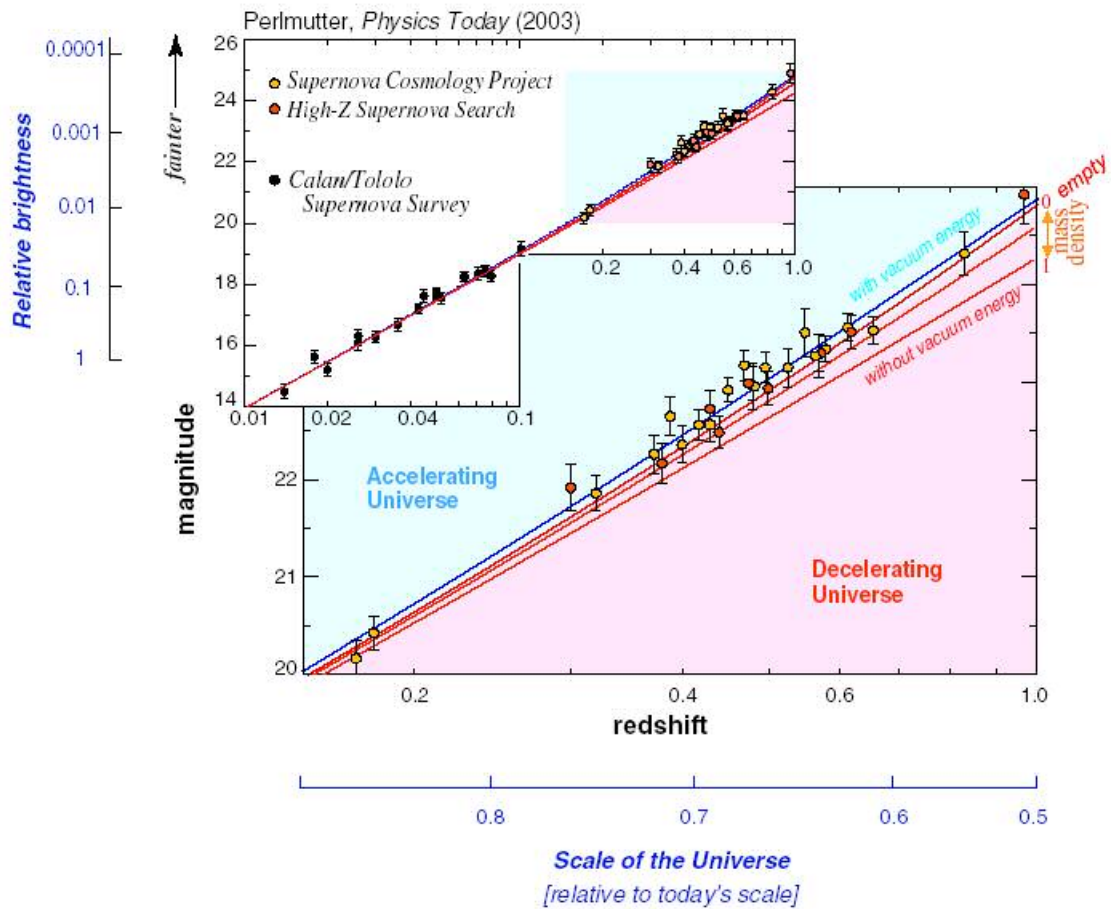


Figure 3: As one looks deeper into space at higher redshifts we see the supernovae, represented as data points, are fainter than we would expect them to be if the universe were expanding at a constant or slowing rate.

Hence, by recent observations, Λ is not just distinctly non-zero, but seems to be playing a non-trivial role in the development of the current universe.

Under this interpretation, the case of a non-zero Λ may be enhanced when the supernovae measurements are combined with measurements of the anisotropies in the CMBR. The temperature anisotropies or temperature fluctuations that we observe are thought to have originated from quantum fluctuations in the early universe. When matter and radiation decoupled ($\sim 300,000$ years after the big bang), photons were then free to

scatter away from their final ‘scattering surface’. It is precisely this scattering effect that Wilson and Penzias observed. Small fluctuations in the radiation energy density at the last scattering surface appear to us as small temperature fluctuations. These anisotropies are one of the benchmarks of modern cosmology. Their origin gives us a hint about the nature of the early universe, and if such temperature fluctuations are indeed caused by quantum fluctuations, then according to those who have come to equate Λ with ρ_{gs} , (the Identity view), it could imply a positive Λ . By the standards of historical researchers, these anisotropies would be a trace of the quantum effects in the early universe and thus support the hypothesis that vacuum energy has been critical in the development of the cosmos. These anisotropies have much potential for further experiments in cosmology (as we will see in section 7). These two types of observations (standard candles and CMBR) offer “cosmic complementarity” in that they work together to support the claim that a non-zero cosmological constant in the form of vacuum energy density may indeed be present in our universe.

Once these observations were made, the status of Λ went from a cosmic accelerator in the early universe as the driving force of inflation that cancelled itself out shortly afterwards, to being a cosmic accelerator in both the early and the *current* universe. The vital point to note here is that the *current* accelerating power of Λ could not have *always* been in effect. Originally Λ was nominated as the driving force behind inflation, but in order for inflation to be a plausible mechanism to be incorporated into our cosmological models, it must only be a very quick phase of acceleration. Inflation suggests a brief period of accelerated expansion followed by a substantial cooling off period so that matter may clump together and form stars and galaxies. If the history of

our universe were one of constant exponential expansion, the outward drive would have prevented the consolidation of matter and we simply would not be here today.

Understandably, the notion of a cosmological parameter that could drive inflation in the beginning of the universe, then be sedated by some internal mechanism that only cancels it to 120 decimal places instead of exactly to zero, while galaxies form, then kicks back into gear to make the universe begin accelerating again does make Λ something of an enigma. Any attempt to resolve the cosmological constant problem would now have to explain the 120-order discrepancy while accounting for these observations.

Section 4.4: The Current Status of the Cosmological Constant and the Problem

Historically, the cosmological constant problem has undergone various changes in its status as a problem. The progression of the changes could be summarized as follows:

- 1) Beginning with the suspicion of Zel'dovich in the late 1960s that zero-point energy could not be ignored when gravity is taken into account led him and others to believe that $\Lambda \neq 0$. In an attempt to 'naturalize' Λ , a discrepancy emerged with observational constraints between 8 and 40+ orders of magnitude.
- 2) The emergence of the Standard Model and electroweak unification in the early 1970s amplified the discrepancy to ~ 120 orders of magnitude and the cosmological consequences of spontaneous symmetry breaking became a relevant feature as to how the problem may be resolved.
- 3) With the advent of inflationary cosmology in the early 1980s, based specifically on vacuum energy to provide the cosmos with the initial boost to instigate exponential expansion, a large vacuum energy density was no longer just a

perplexing notion that needed to be dealt with; it was required to do some explanatory work. This spawned further questions regarding how much vacuum energy is needed for inflation and whether there is any left after the inflationary period is over.

- 4) The observations in the late 1990s indicating that $\sim 70\%$ of the universe is filled with some dark energy source that is causing it to accelerate, put a higher explanatory premium on Λ as a parameter of cosmological significance in both the early and current universe. The role of Λ is now more elaborate and so any attempt to resolve the cosmological constant problem will have to account for this.

The historical narrative above has indicated that Λ emerged in cosmology before it emerged in QFT. Physicists turned to QFT to give a physical basis for Λ that in turn led to the cosmological constant problem. During Λ 's history there have been physicists who have felt no need to bring it into Einstein's equations and others who would think a physical description of the cosmos would be incomplete without it. Among the advocates of Λ , disagreements still arise between those who believe that it must be derivable from the constants of nature in order for it to work as an effective parameter in both QFT and cosmology, and those who may be willing to live with the possibility that no such derivation can be made. These disputes get at the heart of the ontological status of Λ that in turn has made the cosmological constant problem subject to various interpretations. Its status as a 'problem' has changed throughout the years and to this day there is no consensus as to what kind of a problem it is; whether it be a problem of an epistemic or

ontological sort, or not a problem at all. To treat the problem as being epistemic in nature is to presuppose that vacuum energy is the basis of all observations and schematics, and a solution would require giving a consistent account of how it has vanished. To treat the problem as ontological would require that we are certain about what kind of ‘cosmic stuff’ we’re dealing with before an adequate solution could even be posited. And as stated earlier, to treat the problem as a pseudo-problem would be to treat it as something that does not require any sort of solution.

The objective thus far has been solely to lay out the ground work on the subject that has led us to the current state. We may now begin evaluating the variety of solution types that have been posed to solve the problem of Λ . As mentioned in the introduction, the two basic approaches to solving the cosmological constant problem are: the Identity Approach and the Eliminativist approach. Any approach to solving the problem must be able to give an explanation for the inconsistency while at the same time accounting for all the observations. The observations that would need to be explained could be roughly categorized as early and current, namely, accounting for the observed effects of inflation such as spatial flatness and uniformity (from the early universe) and the observed density and acceleration (in the current universe).

Since the inception of the problem in the late 1960s until the observations made in the 1990s much of the work done to solve the problem had been in trying to find some cancellation mechanism to explain how all the vacuum energy disappeared, and if such a mechanism could be adequately described, the problem would be solved. This kind of approach presupposed that the cosmological constant was vacuum energy, $\Lambda \equiv \rho_{gs}$, and so solutions were thought to be found by making theoretical reconstructions involving QFT

in some way. The problem until then was a problem of resolving an inconsistency; vacuum energy was not thought to have any *current* cosmological significance. However, once the recent observations were made, these Identity theorists designated quantum ground state energy as responsible for all of these observed effects. The Eliminativist approach is not committed to such a claim. The Eliminativist approach explicitly rejects the existence of vacuum energy and so is committed to either rejecting all phenomena that vacuum energy is responsible for (either hypothetical or observed) or providing alternative explanations for them. While the Eliminativists need not doubt the recent observations, they attribute another cause to them; instead of the cosmological constant, the observations may be attributable to another dark energy candidate such as Quintessence. If something else is responsible for the universe's missing mass and acceleration, then the cosmological constant could indeed be zero, and accounting for the 120-order inconsistency could be disregarded as an irrelevant puzzle. Of course the Eliminativists would have to be able to tell an effective story as to how inflation occurred and why we need not consider vacuum energy to have any real cosmological significance, but work has been done to further this end, as will be seen in section 6.

The fact that these two approaches to resolving the cosmological constant problem have emerged from the recent observations is significant because it allows physicists to treat the problem experimentally. When asking questions about the universe's characteristics such as the nature of its contents and dynamics, the Identity theorists and the Eliminativists will offer different answers. The distinction between the Identity and Eliminativist views could be stated more precisely in the following equations. Consider, once again, equation (43)

$$\Omega_o^{Total} = \Omega_o^M + \Omega_o^{\Lambda} \approx 1 \quad (43)$$

Where, $\Omega_o^M \approx 0.3$ and $\Omega_o^{\Lambda} \approx 0.7$ and Ω_o^{Λ} is the leftover vacuum energy that's driving the current acceleration. In contrast, an Eliminativist account would modify this equation as

$$\Omega_o^{Total} = \Omega_o^M + \Omega_o^Q \approx 1 \quad (43')$$

Where, $\Omega_o^M \approx 0.3$ and $\Omega_o^Q \approx 0.7$ and Ω_o^Q is the *quintessence* that's driving the current acceleration. The last terms on the RHS of these equations can be tested against one another experimentally, which is to say, the nature of the dark energy that is responsible for the current observations can be determined experimentally. The outcome of such experimentation would be of great assistance to those concerned about the cosmological constant problem, for it would provide them with a more precise direction to field their research. As mentioned, the cosmological constant problem only becomes a problem when we identify Λ as vacuum energy, and the problem becomes more elaborate when this vacuum energy is called upon to do some real work in the current universe. If experimentation can be done to rule out vacuum energy as a dark energy candidate, then Λ is no longer cosmologically relevant and we can dispose of the cosmological constant problem by disposing of the cosmological constant. If experimentation leads to evidence in favor of vacuum energy as the most probable dark energy candidate, then physicists will be much clearer about the work that lies before them for telling a consistent story of Λ . Let us now begin looking at the original efforts that were made to resolve the problem of Λ by way of the Identity approach.

Section 5: The Identity Approach to Solving the Cosmological Constant Problem

In some respects, it is easy to see the appeal of the Identity approach ($\Lambda \equiv \rho_{gs}$). By identifying Λ with ρ_{gs} , we could explain all of the early observations attributable to inflation as well as the current observations as simply being various instantiations of the same kind of cosmic “stuff”. However, by adopting this view we adopt the cosmological constant problem. When Λ was identified as ρ_{gs} it suggested a precursory action to find unity between cosmology and particle physics (or GR and QFT).

The relation between cosmology and particle physics has been expressed as follows: “In a very real sense the job of cosmology is to provide a canvas upon which other fields of science, including particle physics, can weave their individual threads into the tapestry of our understanding of the Universe,” (Kolb, 1994, pg. 362). Kolb proceeds to point out that the aim of modern cosmology is to understand the origin and the large-scale structure of the Universe on the basis of physical law. Furthermore, many physicists today take the ultimate foundation of physical law to be grounded in particle physics. So, is particle physics to be the ‘material’ that covers the cosmological ‘canvas’?

Intuitively, there seems to be a natural connection between cosmology and particle physics. Experiments in particle physics typically involve the study of reaction products from colliding high-energy particle beams where higher energies produce heavier and more exotic particles. Since the Big Bang is believed to have involved extremely high temperatures and energies, with the universe gradually cooling down from a hot and dense initial state, the early universe can be thought of as the ultimate particle physics laboratory. Thus, to study the reaction products of such high-energy

collisions is to get a peak at the conditions of the early universe. In his paper “*Cosmology, particles and the unity of science*,” Zinkernagel (2002) suggests that “Since nobody would deny that nuclear physics is related to particle physics, it would be foolish to assume that the Big Bang nucleosynthesis model is correct but that no particle physics is involved,”³⁵ (pg. 497).

The interplay between particle physics and cosmology originated in work during the 1970s where it was realized that phase transitions associated with field theory might have cosmological consequences (as noted in section 4.2). If cosmology and particle physics are connected as hinted here, there must be some theoretical bridges between descriptions of the very small and the very large, perhaps between GR and QFT, or perhaps even a common framework for both disciplines.

We have already emphasized that the cosmological constant problem is not a problem for QFT or GR in isolation but emerges when these two theories are considered together. So for those concerned with maintaining the Identity view, many different theoretical maneuvers have been suggested, primarily involving the modification of one or both of the theories in some way so that the cosmological constant problem goes away.

Common moves have been to embed either GR or QFT in an extended theoretical framework, such as supersymmetry or quantum gravity in the hopes that it may then be more consistent with its currently conflicting counterpart that would allow the Identity theorists to either formulate a cancellation scheme for the abundance of vacuum energy or give an ‘appropriate’ estimate for ρ_{gs} rather than the enormous $\sim 10^{110} \text{ erg/cm}^3$. Further efforts have been made to dissolve the cosmological constant problem by turning to

³⁵ The nucleosynthesis model explains why we observe such an abundance of light elements in the universe.

semiclassical gravity. These physical frameworks are not incorrect per se and may be fruitful for describing other features of the cosmos, but they run into issues regarding Λ , therefore making it difficult to resolve the issue given the schematics they offer. Since this approach embodies the majority of the research done so far, solutions of this type require some commentary.

Section 5.1: The Supersymmetric Solution

As was seen from the previous section, over the years the cosmological constant problem has changed in light of new cosmological developments and discoveries, and so could be thought of as two issues over this time. The early issue was understanding how the cosmological constant could be zero. The recent issue is now concerned with maintaining a small but non-zero amount of vacuum energy from the vanishing abundance in the early universe in order to account for the recent observations. Supersymmetry was considered as a potential cure to the problem in its early stage.

Although initially investigated for other reasons, supersymmetry (SUSY) was a theory that turned out to be quite relevant to the cosmological constant problem, (for further introductions to SUSY, see Nilles (1984), Lykken (1996), Martin, (1997)). Without diving too deeply into the other physics issues that led to the development of SUSY, it was thought that in QFT there was no limit to how short the wavelengths of a field could be in any particular region of spacetime. If so, an infinite number of different wavelengths within a given region of space could exist, suggesting an infinite amount of ground state energy. Infinite quantities generally tend to be undesirable in physics, and in this case, depending on its equation of state, an infinite amount of energy would

supposedly either send the universe into runaway expansion or produce enough gravitational attraction for the universe to curl up into a single point, which obviously hasn't happened. In the 1970s SUSY was formulated to provide a natural physical mechanism to cancel out these infinities arising from ground state fluctuations. In physics, exactly vanishing entities are generally associated with symmetries. SUSY is a framework within which most models of particle physics can be embedded. It was first formulated in 1971 by P. Ramond while trying to introduce fermions into string theory. He invented an unusual symmetry for exchanging bosons and fermions that came to be known as supersymmetry³⁶.

Generally, particles are classified according to their statistics, namely, integer spin particles are bosons which obey Bose-Einstein statistics while half-integer spin particles, fermions obey the Fermi-Dirac statistics. For instance, photons are bosons while electrons are fermions. Usual symmetries of nature such as Lorentz symmetry do not mix bosons and fermions, they preserve the statistics of each particle. Supersymmetry, however, embeds the Standard Model of electroweak and strong interactions into an extended framework in which each particle has a superpartner. Each superpartner has a spin that is either $\frac{1}{2}$ greater or $\frac{1}{2}$ less than its own, bosons (spin 1, 2, etc) have positive ground states and fermions (spin $\frac{1}{2}$, $\frac{3}{2}$ etc) have negative ground states. So electrons, which are spin $\frac{1}{2}$ particles, have superpartners – the *selectrons*, which are spin 0 particles. Likewise, photons with spin 1 are associated with *photinos* of spin $\frac{1}{2}$. Thus, the *supersymmetric* electroweak theory has twice as many particles as the electro-weak theory.

³⁶ Even though SUSY was formulated amidst the studies of string theory, SUSY is a conceptual scheme independent of string theory.

Another conspicuous feature of supersymmetry is its link to the Poincare' group. The Poincare' group comprises the Lorentz group, i.e. rotations and Lorentz boosts, and the translation group. When transforming a particle into its superpartner and then back into the original particle, the particle is translated in space-time. This fundamental property of supersymmetry is at the origin of its special link with the vacuum energy of the Universe, (See Brax, 2004). In a supersymmetric theory the fermion and boson contributions to the vacuum energy would cancel exactly to zero since they are equally large and have opposite sign. So if we lived in a world in which each particle had a superpartner, we would understand why the vacuum energy vanishes. At least that was the original idea.

For Identity theorists SUSY initially seemed like a good candidate for resolving the cosmological constant problem for it supposedly provided the right kind of cancellation scheme within its theoretical framework such that $\Lambda \approx 0$ would be a sensible result. In a given supersymmetric theory we should be able to calculate the contributions to the energy from vacuum fluctuations. As is hypothesized, in the case of vacuum fluctuations, contributions from bosons are exactly canceled by equal and opposite contributions from fermions – when supersymmetry is *unbroken*. The schematics are relatively similar to other symmetries and dynamics. Just as ordinary symmetries are associated with conserved charges, supersymmetry is associated with “supercharges”, Q , that could be thought of as the ‘generators’ of supersymmetry. These supercharges satisfy the condition that:

$$QQ^\dagger + Q^\dagger Q = 2E \quad (46)$$

in the rest frame of a massive particle of energy E – it is the mathematical formulation of the fact that squaring a supersymmetry transformation amounts to a translation. Taking the expectation value of the left hand side in a given normalized state of the theory, $|\phi\rangle$, gives

$$|Q|\phi\rangle|^2 + |Q^\dagger|\phi\rangle|^2 = 2E \quad (47)$$

from which we deduce that $E \geq 0$ and so the energy is always positive in a supersymmetric theory. Moreover, the supersymmetric vacuum state, $|0\rangle$, such that $Q|0\rangle = Q^\dagger|0\rangle = 0$, has a vanishing energy, $E_{vacuum} = 0$. This is how the mysterious vacuum energy could vanish in a supersymmetric theory. “Such an astonishing result would provide a solution to the cosmological constant problem but for a hitch: the Universe is not supersymmetric,” (Brax, 2004, pg. 232).

One of the initially promising features of SUSY is that it was a cosmological hypothesis that was capable of being tested experimentally. Basically, if every particle has a superpartner, then these superpartners ought to be observable under specific conditions and within specific energy ranges. However, at present, there is no experimental evidence that supersymmetry exists in the real world. No such superpartners have been observed. Thus the above-mentioned cancellation may not have actually taken place, if so the vacuum energy density of the theory ought to be non-zero and quite large. But this absence of evidence need not count as falsifying the SUSY hypothesis. If superpartners do indeed exist, the most likely explanation for why they have not been observed would be that they are too massive to be created in our current particle accelerators. Again, hopefully in the next few years the Large Hadron Collider at CERN

will be ready for use, producing collisions at sufficiently high energies to detect the superpartners many theorists expect to see.

Even if the SUSY particles could be affirmed with experimental data, the role their interactions were to have played in the early universe may not be quite as the Identity theorist would have liked, which has made some physicists incredulous to the powers of SUSY to solve the cosmological constant problem. Brax says, “despite all its promises, supersymmetry alone is not enough to provide a mechanism explaining either the vanishing of the cosmological constant or its extremely small value.” (2004, pg. 233). Others, such as S. Carroll, claim that SUSY is relevant to the problem but is only partially effective; “supersymmetry (SUSY) turns out to have a significant impact on the cosmological constant problem, and may even be said to solve it halfway,” (Carroll, 2000, pg. 26). The rationale behind this latter quote is that since the experimental support for SUSY is still so scarce, at this stage we can only make educated guesses about the energy range where we would expect SUSY breaking to occur and a specific guess would lead to a ‘half-way’ solution to the cosmological constant problem.

If we consider a state where SUSY was broken, we could designate its energy scale as M_{SUSY} . We would then expect a corresponding vacuum energy, $\rho_{vac} \sim M_{SUSY}^4$. In the real world, the fact that accelerator experiments have not discovered superpartners for the known particles of the Standard Model implies that M_{SUSY} is greater than the electro-weak energy scale ($\sim 10^2 \text{ GeV}$), meaning SUSY breaking occurred earlier than electro-weak breaking. As a good guess, M_{SUSY} is at least of the order 10^3 GeV or higher, which is still far greater than what we observe. If we use $(10^3 \text{ GeV})^4$ as a lower limit and compare it with $(M_{vac})^4$ or 10^{-48} GeV^4 , (recall equation 30’), we are left with a discrepancy

$$(M_{SUSY})^4/(M_{vac})^4 \geq (10^{12})/(10^{-48}) \geq 10^{60} \text{ GeV}^4 \quad (48)$$

Obviously this is not a sufficient solution, which may be in part why even more ‘all encompassing’ supersymmetric theories have been constructed, including supersymmetric frameworks that extend into the gravitational sector (supergravity).

Many physicists who believe in SUSY believe its role in alleviating the cosmological constant problem requires the incorporation of gravity into a supersymmetry framework, (Brax, 2004, pg. 232). Brax argues that to date, supergravity schemes have only succeeded in enhancing the cancellation process but not to any degree that could be accepted as a viable solution. In the supergravity framework ordinary particles are complemented with the gravitational multiplet, consisting of the graviton and its superpartner the gravitino. The graviton is the boson that mediates gravity in the same way as photons mediate electromagnetism. The graviton is a spin 2 particle, and its superpartner, the gravitino is a spin 3/2 particle. The graviton and the gravitino are thought to be massless particles when supersymmetry is not broken. But once again, there is no evidence that the world is supersymmetric. In the most common supersymmetry-breaking scheme, the gravitino becomes massive in a manner akin to the way the W boson becomes massive in the electro-weak theory; only in this case it is via the super-Higgs mechanism. As a result, the gravity superpartners become heavier than the particles of the electroweak theory. The mass splitting is of the order of the gravitino mass and is expected, if supersymmetry is ever to be detected, to be of order $10^2 - 10^3 \text{ GeV}$ (i.e. in the electroweak range). A larger mass splitting would rule out supersymmetry as a viable candidate of physics beyond the electro-weak theory, and as soon as supersymmetry is broken, the vacuum energy picks up a value of order

$$\rho_{vac} = (m_{gravitino})^2(m_{Planck})^2 = (10^2)^2(10^{18})^2 = 10^{40} \text{ GeV}^4 \quad (49)$$

Again, this value is a lot less than $\rho_{vac}^{Planck} \approx 10^{72} \text{ GeV}^4$ (recall equation (39)), but still much greater than zero and much higher than the observational constraint. These schematics are what have led many to believe that supersymmetry and supergravity alone are not sufficient to do the desired cancellation work.

Supergravity has endured criticism for reasons beyond its inability to deliver a plausible solution to the cosmological constant problem. Supergravity has encountered problems regarding renormalizability. Originally and up to 1985, most people believed that theories of supergravity would be free of infinities, “then suddenly the fashion changed. People declared there was no reason not to expect infinities in supergravity theories, and this was taken to mean they were fatally flawed,” (Hawking, 2001, pg. 52). Instead it was claimed that SUSY string theory, i.e. Superstring theory, was needed to give an account of SUSY particle physics that was consistent with gravity. Naturally Superstring theory was believed to offer the means of providing a way out of the cosmological constant problem. That is to say, by staying aligned with the Identity view, string theory may provide a mechanism that explains Λ ’s tiny value in a way that SUSY could not. But it is still unknown whether string theory itself offers an accurate description of our actual universe.

Section 5.2: The String Theoretic Solution

Unlike supergravity, string theory initially appeared to be a consistent and well-defined theory of quantum gravity, and therefore calculating a consistent value of the cosmological constant was thought possible, at least in principle. String theory was

originally formulated as a physical model whose fundamental building blocks are one-dimensional extended objects (strings) rather than the zero-dimensional points (particles); it was an effort to avoid the problems associated with the presence of point-like particles in earlier physics. Ripples on the string are interpreted as particles – bosons and fermions. The strings move through a background spacetime, however the spacetime in string theory is not formulated in four dimensions, but in ten or eleven. In order to bring the theory closer to the world we observe, the extra dimensions are “compactified”. Compactification is a process whereby some space dimensions are curled up and are of small size. (For general introductions to string theory, see Green, Schwarz, and Witten (1987) and Polchinski (1998); for cosmological issues, see Lykken (1999), Banks (1999), and Carroll, (2000)).

It was originally thought that for string theory to be consistent, supersymmetry had to be required at some level. In this case, the ground state energies would cancel so exactly that there would be no infinities even of the small sort. If SUSY is preserved, the vacuum energy will remain zero. However, as we know, the results of our SUSY experiments have not been positive, nor do we have reason to believe that the world actually is supersymmetric, and so to describe our actual world we want to break the supersymmetry and be able to account for a small but non-zero quantity of vacuum energy. What followed from the theory is that there could be a large number of possible compactifications, which implied a large number of vacuum states, none of which featured three large spatial dimensions, broken supersymmetry, and a small cosmological constant. The difficulty of achieving a ‘vanishing’ solution to the cosmological constant problem with string theoretic models inspired a number of more elaborate proposals.

By the mid 1980s it was becoming apparent that Superstring theory wasn't the complete picture. Various versions of Superstring theory were emerging that eventually came to be subsumed under the heading of M theory. Even though no one knows what the 'M' stands for, the theory is supposed to be a placeholder for a theory of everything that is to come about in the near future. It is in these more elaborate versions of string theory that physicists were now turning to for a solution to the cosmological constant problem. In the 1990s, after the recent observations were made, Λ was thought to be explicitly non-zero, in which case a clean supersymmetric vanishing scheme was no longer the strategy for solving the problem.

Another proposal put forth was that the absence of manifest SUSY in our world leads us to ask whether the beneficial aspect of canceling contributions to the vacuum energy could be achieved even without a truly supersymmetric theory. In 1999, Kachru, Kumar and Silverstein (1999) constructed such a string theory, and argued that the perturbative contributions to the cosmological constant should vanish to the desired degree. If such a model could be made to work, then perhaps it would be possible that small non-perturbative effects could produce a cosmological constant of a magnitude consistent with astrophysical observations, but this is still a long way off. The actual calculations are somewhat delicate, and the general scheme has not been unanimously accepted (see Harvey, 1999),

Further developments in string theory emerged once it was realized that strings could be thought of as just one member in a wider class of objects called 'branes' that could be extended in more than one dimension. In 'brane' theory, a p -brane has a length in p directions; particles are $p=0$, strings are $p=1$, a membrane or surface is $p=2$, etc. By

thinking in terms of branes, it became possible to set up models of branes being configured in various ways so that the vacuum energy vanishes through dissipation rather than through cancellation.

One popular model is a universe of 11 dimensions consisting of 2 boundary branes with a multidimensional space between them. Our known universe is a 3-brane and functions as one of the boundary branes. The p-number of the other boundary brane has been subject to variation, but is thought to be a 3-brane as well. The remaining dimensions between the boundary branes have been compactified into a Calabi-Yau manifold.³⁷ All the known particles, the strong, weak and electromagnetic forces reside on our 3-brane. Matter and non-gravitational forces like the electric force would be confined to the brane. Thus everything not involving gravity would behave as it would in four dimensions. However, gravity would behave differently from the other forces we experience. In this situation, only gravity is allowed to propagate outside the brane, into ‘the bulk’. The point of the second boundary brane is to prevent gravity from spreading too far into the extra dimensions. In our brane conception of the universe, we would live on one brane but there would be another ‘shadow’ brane nearby. We wouldn’t see this shadow brane because light would be confined to our brane world.

This boundary-brane model is relevant to the cosmological constant problem because the energy created by quantum vacuum fluctuations *may* have a gravitational effect; the gravitational effect of this large amount of energy would then be carried away in the form of gravitons off of our brane and into the bulk. Gravity would permeate

³⁷ A Calabi-Yau manifold is a mathematical construct that describes a multi-dimensional space (first conjectured by Calabi, 1957, and proven by Yau, 1977). It allows for the various dimensions to be compactified while preserving a certain amount of supersymmetry.

through the bulk of the higher-dimensional spacetime and curl it up, which in turn would explain why we don't see the other dimensions and why we don't see very much vacuum energy in the current universe.³⁸ So this would provide an extra-dimensional explanation to the vanishing of the cosmological constant, (for more details see Brax, 2004, and Randall and Sundrum, 1999).

Could it be that the vacuum energy would curve the extra-dimensions while preserving the flatness of our observable Universe? If so, any modification to the vacuum energy, for instance from phase transitions, would not affect the four-dimensional dynamics at all. It would simply modify the geometry of unobservable dimensions. It's a nice idea. The appeal of string theory is that one may incorporate more degrees of freedom in the bulk. "Higher dimensional models provide an interesting framework wherein one may either reformulate or even propose new mechanisms whose four dimensional justification would be debatable," (Brax, 2004, pg. 235).

Even though the string-theoretic schematics do offer some promise to solving the cosmological constant problem, to date the search continues for a (four-dimensional) string theory vacuum with broken supersymmetry and a vanishing cosmological constant. As of today no string theory has made firm predictions that would allow it to be experimentally tested; humans do not currently have the technology to observe strings (which are said to be roughly 10^{-35} cm across). Furthermore, it is not yet known whether string theory is able to describe a universe with the precise collection of forces and matter

³⁸ As the model suggests, by curling up or compactifying 6 of the 7 extra-dimensions on a Calabi-Yau manifold, one gets a five-dimensional model where the extra-dimension is an interval. With compactification, we can make a good approximation of a space-time in fewer dimensions, e.g. one can go from 11 to 5 dimensions by curling up 6 dimensions. The equations of general relativity in five dimensions have solutions where this brane configuration presents flat boundaries, i.e. the metric on the boundaries is invariant under the Poincare' group. This invariance guarantees the vanishing of the cosmological constant in four dimensions as seen by an observer on the boundary branes, (see Brax, 2004).

that we observe. As a theory it might not even be the correct description of nature and its current formulation might not even be directly relevant to the cosmological constant problem. According to Carroll's view of the string theoretic approach to the problem, "It is probably safe to believe that a significant advance in our understanding of fundamental physics will be required before we can demonstrate the existence of a vacuum state with the desired properties. Not to mention the equally important question of why our world is based on such a state, rather than one of the highly supersymmetric states that appear to be perfectly good vacua of string theory," (2000, pg. 30). Ed Witten, a leading expert in the field, shares a similar sentiment regarding the plausibility of string theory to provide a solution to the cosmological constant problem (2000): "As the problem really involves quantum gravity, string theory is the only framework for addressing it, at least with our present state of knowledge ... Assuming that the dynamics gives a unique answer for the vacuum, there will be a unique prediction for the cosmological constant. But that is, at best, a futuristic way of putting things. We are not anywhere near, in practice, to understanding how there would be a unique solution for the dynamics." Unfortunately for the time being, instead of the theory presenting a solution to the puzzle, the puzzle presents a problem for the theory.

Under the presupposition that the cosmological constant is identified as an accumulation of vacuum energy, a new argument has recently been put forth by Rugh and Zinkernagel (2002) to solve the problem in a much different way. Similar in attitude to Ruderman and Bludman (1977), Rugh and Zinkernagel demote the status of the problem to a pseudo-problem by embedding it into a semiclassical framework. They claim that when doing QFT in a semiclassical framework the problem is not even

physically meaningful and so it simply is not a problem to be worried about. To their credit, this conclusion came at the end of extensive research on the problem but is nevertheless not completely convincing. However, there are elements about their view that should not be quickly dismissed, and they provide the basis for taking an Eliminativist approach to the resolving the problem. As experts on the problem, their account is relevant here. They ask us to interpret the cosmological constant problem in a semiclassical way.

Section 5.3: The Semiclassical Solution

First, a few words about the semiclassical approach to QFT. The semiclassical approach to doing QFT entails doing QFT in a curved and dynamic rather than fixed spacetime manifold. To do QFT in curved spacetime is to study the behavior of quantum fields propagating in a *classical* gravitational field. The semiclassical approach is used to analyze phenomena where the quantum nature of fields and the effects of gravitation are both important, but where the quantum nature of gravity itself is assumed not to play a crucial role, so that the gravitation itself can be described by a classical, curved spacetime, as in the framework of GR, (see Wald, 1994). Its applications of greatest interest are to phenomena occurring in the proximity of black holes such as Hawking radiation and events occurring at the early universe such as inflation. The central equation for discussions of quantum field theories in curved spacetime backgrounds is:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = (8\pi G/c^4) \langle \hat{T}_{\mu\nu} \rangle \quad (1')$$

where the notation is the same as in eq.(1), but where the right hand side is now the expectation value of a quantum operator for relevant states such as the ground state.

Equation (1') assumes a specific relation between the *classical* gravitation field and the *quantum* expectation value of $\langle T^{\hat{\mu}\nu} \rangle$, (Rugh and Zinkernagel, 2002).

Rugh and Zinkernagel (2002) claim that, the notion of a cosmological constant problem rests fundamentally on two assumptions – both of which can be questioned:

- (i) The semiclassical approach, in which quantum energy acts as a cosmological constant, is valid
- (ii) The QFT vacuum energy is physically real (as in the standard QFT interpretation)

Rugh and Zinkernagel take issue with both of these assumptions. If these assumptions are indeed problematic then that would have a bearing on how we are to interpret the problem with the cosmological constant. Let us look at these two assumptions, the issues that they have with them, and see what this means for the cosmological constant problem.

Assuming the Validity of the Semiclassical Approach

Note as the first of these assumptions is presented – ‘in which quantum energy acts as a cosmological constant’ – implies $\Lambda \equiv \rho_{gs}$. Rugh and Zinkernagel claim that when designating vacuum energy as part of the matter-energy content of universe and playing a significant role, which the Identity theorists are committed to, the semiclassical formulation is problematic both at a technical and at a conceptual level (and this is of course with the second assumption in place that vacuum energy is physically real).

Technically, it is difficult to calculate $\langle T^{\hat{\mu}\nu} \rangle$ in a curved spacetime³⁹ and conceptually, in

³⁹ Rugh and Zinkernagel claim that a problem with equation (1') is the ‘back-reaction’ problem that would need to be resolved in some self-consistent way if the semiclassical approach is to survive. In essence, if

a curved spacetime, Poincare invariance is lost and so the vacuum state cannot be well defined in an observer-independent way. The Poincare invariance of the vacuum state in Minkowski space implies that the vacuum expectation value of the stress energy tensor, i.e. $\langle 0 | T_{\mu\nu} | 0 \rangle$, is Poincare invariant. In a space curved by gravity, “the gravitational field (e.g. in our expanding universe) will in general be expected to produce particles, thereby obscuring the concept of vacuum as a state with no particles ... so whereas the state $|0\rangle$ with no particles in it is the obvious vacuum state in Minkowski spacetime, in general there is not any reference system in which there is no particle production. With no clear vacuum concept, one cannot give a precise meaning to $\langle 0 | T_{\mu\nu} | 0 \rangle$, let alone associate this quantity with a cosmological term,” (Rugh and Zinkernagel, 2002, pg. 691).

According to Rugh and Zinkernagel, if we were to believe that $\Lambda \equiv \rho_{vac}$, and we believe the semiclassical approach is valid, and within the semiclassical approach a distinct value for ρ_{vac} cannot be determined, then there’s no theoretical basis for the inconsistency of Λ . If the problem rests on a semiclassical approach to doing QFT, and within the semiclassical approach the cosmological constant problem cannot even be meaningfully posed, then there is no problem. Formally speaking, the argument that Rugh and Zinkernagel have advanced is represented as follows:

- 1) The semiclassical approach is valid.
- 2) The cosmological constant problem requires a distinct value for vacuum energy density in order for the problem to be meaningfully posed.

$\langle T_{\mu\nu} \rangle$ affects the metric, then this metric will change the assumptions for calculating $\langle T_{\mu\nu} \rangle$ (see Wald, 1994, p.54). Solutions to this ‘chicken and egg’ issue have been put forth by Calzetta and Hu (2005).

- 3) The semiclassical approach implies no preferred vacuum state.
- 4) No preferred vacuum state implies no distinct value for the vacuum energy density.
- 5) Therefore, the cosmological constant problem is not well formulated and is not a physically meaningful problem.

Should we take their argument seriously, then the problem would be essentially *dissolved*, and we could all begin working on something else. However, people are still working on the problem. There are some relevant reasons for this. The first is that Rugh and Zinkernagel are suggesting that because the cosmological constant cannot be effectively posed *semiclassically*, then it's no longer a problem. To be fair, they only suggest this as a stance one can take toward the problem, but are not explicitly endorsing it. In order to accept the cosmological constant problem as sufficiently dissolved in this way would require showing that all other approaches taken to solving it can all be subsumed under the semiclassical umbrella. It is not to say that such a feat is impossible, but it has yet to happen and would take some work. At present, it seems like a leap to go from admitting that the argument cannot be posed semiclassically, to admitting that the argument is un-posable or does not exist, but then again, if we take the semiclassical approach as valid and sufficiently applicable to the structure of the actual universe, then perhaps we could argue that no real solution can be posed because there is no real problem in our midst. There is certainly an appeal to this line of reasoning, but the argument that Rugh and Zinkernagel have presented to us is not completely persuasive as

is, and may enable those who are really concerned about the cosmological constant problem to look beyond it.

As it stands Rugh and Zinkernagel's argument scheme is based on a problematic premise. Their argument is based on the premise that the cosmological constant problem requires a distinct and preferred value for vacuum energy density (premise 2), and such a value cannot be obtained when doing QFT in a curved spacetime. However, not everyone would agree with this point. In his book, Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics, (1994) Wald, claims that in a curved spacetime the vacuum state *can* be defined in an observer-independent way. Wald would be willing to admit that Poincare invariance is lost in curved spacetimes, which makes it impossible to speak about a preferred state of no particles, but he does not take this to be a serious issue. In Wald's approach to semiclassical QFT he de-emphasizes the role of particles. He states, "as a matter of principle, it should be stressed that, in general, the notion of "particles" in curved spacetime is, at best, an approximate one" (pg. 59). Wald views the "the lack of a preferred notion of "particles" in quantum field theory in curved spacetime to be closely analogous to no preferred system of coordinates in classical general relativity ... in both cases the lack of an algorithm does not, by itself, pose any difficulty for the formulation of the theory," (pg. 60). As an alternative, he suggests an algebraic approach to defining the vacuum state in an observer-independent way, (for further details, see Wald, 1994, chapter 4). The notable point is that the loss of Poincare invariance does not necessarily make the vacuum state indeterminable.

Also, Earman (2003) has raised a concern about the cosmological constant problem even if we accept Rugh and Zinkernagel's point about the loss of Poincare invariance in

curved spacetime leading to no preferred vacuum state. He considers the case where the stress-energy tensor of quantum fields, $T^{\hat{\mu}\nu}$, arises from a scalar field Φ . In a regime where the field Φ is not changing, its potential would function as an effective positive cosmological constant. Thus, the questionable condition of Poincare invariance does not have to be invoked for a surrogate cosmological constant to emerge. The quantum expectation value of Λ^{Φ} , whether computed from some preferred vacuum state or not, must be consistent with observational limits on Λ , (Earman, 2003, pg. 567).

Even if the inability to identify a preferred vacuum state in a curved spacetime were a significant issue, and even if the semiclassical approach is indeed valid, the cosmological constant problem is still a posable problem given the conditions of our actual universe. From observations in astrophysics it is known that the gravitational field in our local neighborhood (our solar system etc.) is rather weak, i.e. in suitably chosen coordinates the spacetime around us may be written as the flat Minkowski spacetime metric plus a small perturbation from weak gravitational fields. We have also seen that under the influence of small gravitational perturbations e.g. due to the earth's gravitational field the predictions of QFT such as the Cassimir effect and the Lamb shift⁴⁰ remain stable and these effects can still be observed. Thus, it is reasonable to expect that we can accurately apply standard QFT in our local astrophysical neighborhood, where the perturbations due to gravity are not sufficient to curve the spacetime enough to make it impossible to speak meaningfully about the vacuum state. Furthermore, with respect to the universe at large, astrophysical evidence is consistent with cosmological models where the expansion is rather slow and the metric close to spatially flat. Under these

⁴⁰ A split of energy levels in the hydrogen atom.

circumstances, it may be possible to define an approximate vacuum state of the quantum fields in the universe that has almost no particles in it, (See Birrell and Davies, 1982, pg. 73). Even if we take the semiclassical approach as valid and consistent with the actual universe, it remains the case that the actual universe can be treated as slowly expanding and basically flat (due to the overall low density of matter) to the extent that the problems of doing QFT in curved spacetime can be ignored.

As we have seen that contrary to Rugh and Zinkernagel's concerns, the cosmological constant problem is still a problem when we interpret it semiclassically. And even if the issues that arise from doing QFT in a curved spacetime did render the problem ill-defined, we could treat the spacetime in our universe as basically flat, in which case the cosmological constant problem could still be meaningfully posed. If their issues regarding the semiclassical approach are not sufficient to justifiably throw out the problem, it raises the question of how the semiclassical approach could solve the problem.

The semiclassical approach is not any more conducive to the Identity approach than to the Eliminativist approach or any other approach, but in this section I will show how it could be used to formulate a solution that would be satisfactory to the Identity theorists. As we'll see the semiclassical approach suggests the possibility that both positive and negative energy modes may exist in the quantum vacuum, and the solution would involve a cancellation between these positive and negative energy modes leaving the total value of $\rho_{vac} \approx 0$. But let's review some of the key features of semiclassical gravity that have been critical in the development of explanatory schemes relating to

other cosmological phenomena, namely, the inflationary epoch in the early universe and Hawking radiation in the proximity of black holes.

The inflationary epoch is commonly thought to have been instigated by quantum vacuum energy, and in the early universe when the inflationary epoch occurred, the matter and energy were much more condensed and spacetime was much more curved than today's roughly flat spacetime. Hence, many inflationary models are semiclassical in kind since inflation is thought to have resulted from quantum field theoretic effects in a curved spacetime. Even though there are many different kinds of inflationary models, a large set of them tend to treat the matter in the universe as a perfect fluid. For these models to do the work we want them to do requires that this fluid has a special equation of state, $p = -\rho$, where p and ρ are its pressure and density, respectively. With an equation of state involving negative energy density, this negative energy (often attributed to a scalar 'inflaton' field) produces a repulsive gravitational effect that drives the exponential expansion. Inflationary models are attractive in cosmology because they explain various issues that could not be explained in the standard big bang framework, such as: the horizon problem, the flatness problem, and the monopole problem. However, in order for inflationary models to do the work we want them to do, they must violate all of the standard energy conditions that regular matter (the contents of $\langle \hat{T}_{\mu\nu} \rangle$) is to comply with⁴¹. In essence, the energy conditions imply that all the energy that can be related to matter must be non-negative, $\rho \geq 0$.

⁴¹ The *Strong* energy condition ($\rho + 3p > 0$ and $\rho + p > 0$)
 $(T_{ab} - 1/2 g_{ab}T) V^a V^b > 0$ for all timelike V^a , where T is the trace of T_{ab} .
The *Dominant* energy condition ($\rho > 0$ and $\rho \pm p > 0$)
Requires that $T_{ab}V^a$ is non-spacelike for all timelike or null V^a .
The *Weak* energy condition ($\rho > 0$; $\rho + p > 0$)
Requires that $T_{ab}V^a V^b > 0$ for all timelike V^a .

Negative energy also plays a role in the other semiclassical context of black hole evaporation and Hawking radiation. In the early 1970s, in an effort to show that it's possible for black holes to emit radiation and eventually evaporate, Stephen Hawking presented a semiclassical argument, although he did not refer to it in that way specifically. In the powerful gravitational field surrounding the black hole, virtual particles pairs are created in the same way that virtual particles are created in the quantum vacuum. These virtual pairs contain energy of opposite signs. They are created just beyond the event horizon of the black hole where subsequently, some particles travel into the black hole while some particles travel off into space. Generally speaking, particles with positive energy that fall into a black hole will contribute to its mass, and particles with negative energy will deplete its mass. The negative particles from the pair production that fall in are responsible for the black hole's 'evaporation' and the positive particles traveling outward would be the detectable radiation of this phenomenon.

The two most salient examples of cosmological events that are explained semiclassically, inflation and black hole evaporation, both violate the standard energy conditions of regular matter and not only allow for the scalar fields with negative energy in the universe but require it. An Identity theorist intent on resolving the cosmological constant problem semiclassically could similarly rely on negative energy to aid in his/her efforts. In a way similar to how the SUSY theorists relied on a cancellation scheme between all particles and their super-partners to explain why the value of Λ is zero instead of the huge quantity that QFT predicts, a semiclassical theorist could invoke a

cancellation scheme between positive and negative energy fields. Consider once again equation (36)

$$\rho_{vac} = E/V = (1/V) \sum_k \frac{1}{2} \hbar \omega_k$$

and equation (39)

$$\rho_{vac}^{Planck} \approx (10^{18} \text{ GeV})^4 \approx 10^{72} \text{ GeV}^4 \approx 10^{110} \text{ erg/cm}^3.$$

The large value for ρ_{vac} (equation 39) is obtained by treating the EM field as a set of harmonic oscillators and summing over all the zero-point energies for each oscillator mode (equation 36). For these quantum field theoretic equations to obtain their present value, an assumption is made that all of the oscillator modes are positive, i.e. $\frac{1}{2} \hbar \omega_k \geq 0$. However, if we take the notion of negative energy seriously, as the semiclassical approach to QFT suggests, there would be negative oscillator modes as well, $\frac{1}{2} \hbar \omega_k \leq 0$. Hence, equation (36) could be reformulated as:

$$\rho_{vac} = E/V = (1/V) \sum_k (\frac{1}{2} \hbar \omega_k) + (1/V) \sum_k (-\frac{1}{2} \hbar \omega_k) \quad (36')$$

When we sum up all the oscillator states, some will be positive while others will be negative, and could either cancel to zero or achieve a result that is consistent with observational constraints. At present, the semiclassical approach to QFT cannot say specifically why Λ or ρ_{vac} has the small yet non-zero value it has, but it can nevertheless provide the Identity theorists with solid grounds for why there should not be the problem of 120 orders of inconsistency. To ensure that scalar fields, both positive and negative, are generated and function in the way that can effectively explain the relevant phenomena, they must be subject to certain constraints and could only take on certain values. What these values are is still a matter of speculation but work is currently being done to determine what the lower bounds may be (see Calzetta and Hu, 2005). If the

lower energy bound could be appropriately estimated semiclassically, then the semiclassical approach would have theoretical grounds for why we should expect to see the value of Λ that we see and not worry about the abundance of missing energy. But like the previous Identity schemes of SUSY and string theory, this is still a promissory note.

From the commentary on the first assumption we see that despite Rugh and Zinkernagel's concern, when we consider the semiclassical approach, in which vacuum energy acts as a cosmological constant, to be valid, the cosmological constant problem can still be posed and semiclassical solutions can be presented (albeit promissory). But these solutions, as well as all the other Identity-based solutions, assume the second assumption that vacuum energy is physically real. However, the second assumption regarding the existence of vacuum energy, if it turned out not to exist, would have significant consequences regarding how we are to interpret the problem. Would there still be a problem? Let us now consider this second assumption.

Assuming the Reality of Quantum Vacuum Energy

As early as the 1930s, the question of how accurately one can measure the components of the quantized electromagnetic field was addressed by Bohr and Rosenfeld, (See Bohr and Rosenfeld, 1933). Bohr claimed, "The idea that the field concept has to be used with great care lies also close at hand when we remember that all field effects in the last resort can only be observed through their effects on matter. Thus, as Rosenfeld and I showed, it is quite impossible to decide whether the field fluctuations are already present in empty space or only created by the test bodies," (Bohr in a letter to Pauli, 1934, as appears in Rugh and Zinkernagel, 2002). They argued that the QED formalism reflects an

unphysical idealization in which field quantities are taken to be defined at definite spacetime points. But when measuring the field strengths, only an ambiguous meaning can be attached to average values of field components over a finite spacetime region. In the simplest realization of such measurements, an arrangement of charged test bodies is envisaged in which the test charges are homogeneously distributed over the finite spacetime region. In their expression for the measurement uncertainty of the field strengths, Bohr and Rosenfeld showed that these diverge when the volume of the spacetime region approaches zero. This means that field fluctuations are ill-defined at definite spacetime points. Bohr and Rosenfeld concluded that the origin of field fluctuations in a measurement arrangement is unclear, since the result of a measurement of the field fluctuations rests on the charged test bodies in the finite spacetime region.

The ambiguity regarding the origin of the field fluctuations translates into an ambiguity regarding the origin of the vacuum energy, and so what we measure as vacuum energy *might* be the result of an experimental arrangement rather than a feature of the vacuum ‘in itself’. As is known, since Bohr and Rosenfeld’s work, several experiments have been conducted to observe the physical effects of the vacuum, such as the Casimir effect and the Lamb shift. The important point is that the measurements are on material systems such as the plates in the Casimir effect, the atom in the case of the Lamb shift, etc. From these experiments it seems impossible to determine whether the experimental results point to features of the vacuum ‘in itself’ or of the material systems trying to measure it. Rugh and Zinkernagel claim “Insofar as there are no clear experimental demonstrations of the reality of vacuum energy in the laboratory, one could also speculate that there is no real vacuum energy associated with any of the fields in QFT,”

(2002 pg. 685). Certainly if vacuum energy did not exist, then speaking about a 120 order inconsistency with respect to a fictional entity would not be a problem we would need to concern ourselves with. But can we dismiss the reality of vacuum energy so quickly? It would surely clear up the problem of Λ , but we would be left wanting explanations for all events supposedly caused by vacuum fluctuations and energy, such as black hole radiation, inflation, cosmic acceleration, etc. Also, what would the dismissal of ρ_{vac} mean for QFT in general?

Rugh and Zinkernagel speak of the unreality of ρ_{vac} only speculatively, but they do not wish for such speculation to necessarily imply that the standard QFT formalism is altogether misleading. They offer a possible interpretation of QFT should we believe that there is no real vacuum energy. Should we choose to take an Eliminativist stance toward vacuum energy then we can maintain the standard QFT formalism but one should not associate energy with fields in empty space. The vacuum energy is therefore to be viewed as a mathematical artifact of the theory with no independent physical existence. This view is consistent with the fact that vacuum energy can be a practical concept in connection with deriving quantum features of material systems such as the Casimir effect – insofar as these features can also be accounted for by referring to the material constituents of the systems studied. Another possible interpretation is that the standard QFT formalism is abandoned altogether and replaced by something like Schwinger's source theory. According to source theory, there are no quantum operator fields and there are fields only when there are sources (e.g. the material constituents of the Casimir plates). What this means in particular, as emphasized by Schwinger, is that there is no energy in empty space: "...the vacuum is not only the state of minimum energy, it is the

state of *zero* energy, zero momentum, zero angular momentum, zero charge, zero whatever.” (See Schwinger, 1973). Similarly, under this scheme, quantum fluctuations in empty space would be non-existent and so all phenomena attributed to them, either hypothetical (such as Hawking radiation) or observed (such as cosmic acceleration) would have to be rejected or explained through alternative means.

Rugh and Zinkernagel’s concern regarding the reality of vacuum energy is a legitimate one that would force us to reconsider the status of the cosmological constant problem if it did not actually exist. After all, how could there be a cosmological constant problem if that which we identify as the cosmological constant, ρ_{vac} , does not exist? However, simply rejecting ρ_{vac} does not dissolve the problem. One cannot simply reject ρ_{vac} without offering an alternative explanation for the relevant phenomena.

For the Eliminativists, the uncertainty that emerges from the observations of vacuum energy (Cassimir effect and Lamb shift) raises the question of whether the substantial conception of a QFT vacuum with non-zero energy in ‘empty space’ (existing in the absence of any material constituents) ought to be the basis for such significant cosmological events such as inflation and the current acceleration. It is precisely because of the cosmological premium that has been placed on ρ_{vac} that the cosmological constant problem has become one of the uglier monsters of modern physics. Those who accept that ρ_{vac} exists and was very large in the early universe are tasked with explaining how such a large quantity has mysteriously cancelled itself out to 120 decimal places over the past 14 billion years to accord with modern observations. While schemes to this end have been established and even show promise, they have also shown to be still problematic. Perhaps a simpler solution to the cosmological constant problem may exist by merely

eliminating the cosmological efficacy of ρ_{vac} all together, provided that alternative explanations can be given to account for the relevant observations.

As we'll see in the following sections, a benefit to taking the Eliminativist approach seriously as a viable alternative to the Identity approach is that it opens up experimental possibilities for testing these approaches against each other. And through experimentation, progress can be made toward finding a plausible and workable solution to the problem.

Section 6: The Eliminativist Approach to Solving the Cosmological Constant Problem

As mentioned in section 3, Λ has always had its champions such as Lemaitre and Eddington. Both Lemaitre and Eddington thought Λ was necessary to explain various cosmological phenomena such as cosmic expansion and structure formation. These phenomena do warrant explanations. To reject Λ outright and offer no explanation would be to neglect their concerns, but we can satisfactorily account for these phenomena by appealing to parameters other than ρ_{gs} . By taking this route, it could satisfy the explanatory demands of the one-time champions of Λ , as well as avoid the (120 order) inconsistency by sending it back to the shelf again. In essence, to eliminate the cosmological constant is to eliminate the cosmological constant problem. Instead, the need to bring Λ into Einstein's field equations and the need to attribute so much cosmological significance to quantum vacuum fluctuations is eliminated by calling upon other sources of energy to account for inflation and some other dark energy candidate to account for the recent observations⁴². Earman has noted, "Astrophysicists who are unable to swallow Λ have postulated a new hypothetical form of matter called "quintessence" that will mimic some of the key effects of Λ ," (2001, pg. 214). These 'key effects' are essentially an observable, smoothly distributed abundance of energy throughout the cosmos that is responsible for its expansion and acceleration.

Distinct from Λ 's status as a constant quantity of unknown (or dark) energy in Einstein's field equations, these 'Quintessence' models treat the dark energy component

⁴² At the same time, resolving the cosmological constant problem in this Eliminativist way would entail the denial of all hypothetical phenomena that are thought to occur as a result of vacuum fluctuations such as Hawking radiation and black hole evaporation. At present there have been no observations to support such phenomena, which is consistent with the Eliminativist account of ρ_{gs} but should such evidence ever appear the burden would be on the Eliminativist to account for it without mention of vacuum fluctuations.

as dynamic rather than constant⁴³. In his paper, “*The Cosmological Constant*”, Carroll (2000) claims, “There are many reasons to consider dynamical dark energy as an alternative to a cosmological constant. First and foremost, it is a logical possibility, which might be correct, and can be constrained by observation. Secondly, it is consistent with the hope that the ultimate vacuum energy might actually be zero ... But most interestingly, one might wonder whether replacing a constant parameter Λ with a dynamical field could allow us to relieve some of the burden of fine-tuning that inevitably accompanies the cosmological constant,” (pg. 35).

Any fundamental theory of nature that intends to successfully generate a Λ -term will be confronted by a fine-tuning problem for two reasons. One is that during the expansion of the universe the energy density of matter decreases as a^{-3} while the density of ρ_{gs} remains constant. As a result, in a universe of broken symmetry a significant fine-tuning of initial conditions would be required in order to ensure that the cosmological Λ -term comes to dominate the expansion dynamics of the universe at precisely the current epoch, no sooner and no later. Also, some fine-tuning would seem necessary somewhere to account for why the Λ -term is so miniscule compared with its predicted value at either the Planck or the electroweak scale. The fine-tuning problem is rendered less acute if we relax the condition that the energy density that Λ represents must be constant, and try to construct models where the energy field is dynamic, (see Sahni, 2002). Substituting the enigmatic ρ_{gs} that Λ was thought to be with a new dynamic substance is certainly possible given current cosmological models, but can Quintessence play the same explanatory role that Λ was thought to play? In this chapter we will look at the Eliminativist accounts of

⁴³ The Quintessence scenario amounts to a semiclassical approach in that Quintessence rather than ρ_{gs} is coupled to Einstein’s field equations.

the recent observations and of inflation, respectively. Let us begin with some general remarks about how these Quintessence models are supposed to function.

Section 6.1: The Essence of Quintessence

The Rise of Dark Energy

As with many of the techniques employed for solving the cosmological constant problem, the original impetus to develop Quintessence models was not to resolve the problem, but such models have since come to be seen as an alternative to the original Λ . In the last decade, various discoveries (noted in section 4.3) caused many cosmologists to reconsider their beliefs about the composition and behavior of the universe. New features were being discovered that needed to be accommodated. Since all the matter in the cosmos (both baryonic and dark) only adds up to roughly 30% of the critical density (see section 4.3), accounting for the missing 70% became a new puzzle for cosmologists. And as the observations of Type 1a supernovae began to accumulate, these observations indicated that the universe is not only expanding but also accelerating.

To some cosmologists, these new features were more difficult to explain in terms of some leftover quantity of vacuum energy. Again, maybe there is an elusive cancellation mechanism that instead of canceling Λ (or ρ_{gs}) to exactly zero, it only cancels to 120 decimal places. Then, the leftover vacuum energy would comprise the missing 70% of the critical density, but a conceptual difficulty arises when we consider the state of the current acceleration. In his paper, “*Quintessence*” (2000), R. Caldwell notes the crucial point that the *accelerating* power of Λ could not have *always* been in effect. Λ was originally nominated as the driving force behind inflation, but in order for

inflation to be a plausible mechanism to be incorporated into our cosmological models, it must only be a very quick phase of acceleration. Inflation suggests a brief period of accelerated expansion followed by a substantial cooling off period where matter can clump together and form the large-scale structures of the universe. If the history of our universe were one of constant exponential expansion, the outward drive would have prevented the consolidation of matter and we simply would not be here today.

Understandably, the notion of a cosmological parameter that could drive inflation in the beginning of the universe, then be sedated by some unknown, internal mechanism while galaxies form, then kicks the universe into accelerated expansion once again, prompted some cosmologists to explore other possible explanations, at least for the current observations. As has been suggested, the Eliminativist approach entails giving explanations to the recent observations and to account for inflation other than appealing to vacuum energy.

Although a cosmological constant is a respectable fit to the current data, so is Quintessence. As plausible as both explanations are, these two dark energy candidates yield slightly different observational consequences and so are subject to experimental tests, as we'll see in the next section. Basically, whatever dark energy is exactly, it is thought of as a fluid with sufficiently negative pressure and a negative equation of state that will cause the expansion to accelerate. Cosmic acceleration can only occur if the pressure is sufficiently negative. This is because general relativity tells us that energy and momentum, and therefore pressure, all gravitate. The strength of this gravitational force is determined by the equation of state, $\rho + 3P$. If $\rho + 3P$ is negative, which can happen for negative pressures, then the gravitational force is repulsive. When we apply this feature

to the universe, the effect is a ubiquitous energy substance with negative pressure causing space to repel itself – every point in space diverges from its neighbors and the cosmic expansion accelerates.

Aside from accounting for ~70% of the energy density in the universe today, dark energy must have been an insignificant fraction earlier in the universe, otherwise its influence would have made it almost impossible for ordinary matter to form the stars, galaxies and large-scale structure that we see in the current universe. In this case, any form of energy that dominates today, but was insignificant in the recent past, must have a density that decreases much more slowly with time than the matter density. That is, as the expanding universe doubles in volume and the matter density decreases by a factor of two, the density of this dark energy must decrease by a smaller factor. A kind of dark energy with this feature can be accommodated by Einstein's equations (equation 2) to account for the 'accelerating' observations provided it has a negative equation of state and is then gravitationally self-repulsive. This is the feature of dark energy that does not mesh well with our conception of Λ as vacuum energy density since up to the present, the Identity theorists who believe in the cosmological significance of vacuum energy have considered it to be of constant density⁴⁴.

In Quintessence models where dark energy is treated dynamically, its density starts out as similar to the density of matter and radiation in the early universe. The Quintessence and matter density would both decrease at similar rates as the universe

⁴⁴ It should be noted that there's nothing about the Identity approach that requires a conception of vacuum energy that is constant. Arguments could be made that the vacuum energy density may increase as the universe expands insofar as a greater volume allows for energy modes that could not exist in smaller volumes. Hence, a greater number of energy modes per unit volume could increase as the universe expands. However, not much work has been done to this end and as has actually been the case, all Identity theorists have considered ρ_{gs} to be a constant quantity.

expanded, but the quintessence density would overtake the matter density once structure has formed in the universe. By treating Quintessence dynamically it can account for how it gradually develops negative pressure sufficient enough to overcome the attractive gravitational force of all the energy density in the universe and to drive the accelerating expansion, (for further reading, see Turner and White (1997), Carroll (1998), Cornish and Starkman (1998), Huey et al. (1999), Spergal and Steinhardt (1999), Zlatev et al. (1999), Zlatev and Steinhardt (1999), Wang et al. (1999)).

On the Quintessence scenario, the dynamical constraint is $\Omega^{Mc} \leq 0.3$, where M_c stands for ordinary gravitationally clustering matter, allowing for the possibility of Quintessential matter (or energy) M^Q to make up all of the missing mass required by the standard inflationary scenario. Thus

$$\Omega^{total} = \Omega^{Mc} + \Omega^{MQ} \approx 1 \quad (50)$$

without any help from Λ , (this is a modified version of equation (43), that is (43')). The idea of Quintessence could be thought of as a single and uniform scalar field, $\phi(t)$, that slowly rolls down a potential $V(\phi)$, at a rate governed by the field equation (Weinberg, 2000):

$$d^2\phi/dt^2 + 3Hd\phi + V'(\phi) = 0, \quad (51)$$

where H is the Hubble parameter governing the expansion rate, represented here as

$$H = \sqrt{(3/8 \pi G) \cdot (\rho_\phi + \rho_M)}. \quad (52)$$

Here ρ_ϕ is the energy density of the scalar field, represented as the sum of its kinetic and potential energy:

$$\rho_\phi = \frac{1}{2} d\phi^2 + V(\phi) \quad (53)$$

while ρ_M is the energy density of matter and radiation, which decreases as

$$d\rho_M = -3H (\rho_\phi + p_M) \quad (54)$$

with p_M being the pressure of matter and radiation.

If there is some value of ϕ where $V'(\phi) = 0$, then given the dynamic nature of ϕ , it should approach this value, but only *slowly* with time. Meanwhile ρ_M is slowly decreasing, so that eventually the universe starts an exponential expansion with a slowly varying expansion rate $H = \sqrt{(8\pi G \cdot V(\phi)/3)}$. To achieve this effect, quintessence mimics a perfect fluid. Its pressure is identified as the difference between its kinetic energy and potential energy and the energy density is the sum of these energies. It takes on an equation of state as:

$$P_Q = w_Q \rho_Q \quad (55)$$

Whose equation of state parameter, w_Q , is

$$w_Q = p_\phi/\rho_\phi = (\frac{1}{2} d\phi^2 - V(\phi)) / (\frac{1}{2} d\phi^2 + V(\phi)) \quad (56)$$

Thus, when the field is slowly varying and $d\phi^2 \ll V(\phi)$, we have an equation of state where $w \geq -1$ and the scalar field potential acts like a cosmological constant, (see Carroll, 2000). For properly chosen potentials the value of the field and its kinetic energy can be arranged in such a way that the energy density is $\sim 70\%$ of the critical energy density of the universe now. In other words, properly chosen potentials can account for the ‘missing mass’ problem in a manner as effective as Λ is supposed to.

The observations indicate that the energy responsible for them has an equation of state, where $w \geq -1$. In a Λ model, since ρ_{gs} is constant, w is always ≥ -1 . Again this requires fine-tuning to set the constant value and to ensure the dynamics its responsible for occur when they are supposed to occur. By allowing dark energy to be time-

dependent we can obtain a similar equation of state by choosing a dynamic potential. Quintessence offers a dynamical account of why $w \geq -1$. By invoking a dynamical scalar potential, as we'll see in the Quintessence models presented below, we can arrive at an acceptable value for the energy density necessary to mimic the effects of Λ without any cancellation mechanisms or fine-tuning. These Quintessence models are especially attractive considering a lingering issue that cosmologists still find difficult to explain, namely, why the acceleration should begin at this particular moment in cosmic history. Is it a coincidence that just when thinking beings have evolved the universe suddenly shifts into overdrive? Without invoking any anthropic explanations⁴⁵, perhaps a more satisfying possibility is that the acceleration is triggered by natural events in the later history of the universe. For instance, once the “matter-dominated epoch” was underway and gravity could begin to clump matter together to form stars, galaxies and large-scale structure, is it possible that this transition triggered the onset of Quintessence? A set of ‘tracker’ Quintessence models have been put forth that describe this exact state of affairs.

Tracker Models of Quintessence

To date, investigations have focused on tracker models of quintessence where the scalar field energy density evolves in parallel with that of matter or radiation, at least for part of its history (see Ferreira and Joyce (1998), Copeland et al. (1998)). As previously stated, in the early universe the dark energy density is thought to be comparable with the matter density. The Quintessence, “tracks” the radiation and matter density. When the universe is radiation dominated, the tracker field mimics radiation: the energy density

⁴⁵ See Appendix for details

falls at the same rate as the radiation energy density, and its equation of state is given by $P = \rho/3$, the same as for radiation. When the universe becomes matter dominated, the tracker field mimics the matter, for which the pressure is nearly zero.⁴⁶ As long as the mimicking continues, the tracker-field energy is a small fixed fraction of the total energy and the expansion is constant (or possibly even decelerating). For most of the history of the universe, the Quintessence occupies a very small fraction of the critical density, but the fraction grows slowly until it catches up with and ultimately overtakes the matter density causing a period of acceleration.

If the tracker field is constant, the kinetic energy is negligible compared with the potential energy, which is precisely the condition required for a negative pressure component and cosmic acceleration. The tracker potential must possess some feature that causes the field to become locked into a nearly constant value at some later time. The problem is that the feature of the potential that locks the tracker field must be delicately tuned so that the acceleration begins at the right time.

This fine-tuning feature can be circumvented by adopting a novel form of tracker field called "k-essence", short for kinetic-energy-driven quintessence (see Armendariz et al., 2000). Original Quintessence models allowed for dark energy density to be time-dependent, but still maintained a constant non-positive equation of state, and explaining this cosmic coincidence required choosing the proper value for the potential. In these k-essence models, not only is the dark-energy density time dependent, but so is the equation of state. In these models, the kinetic energy depends nonlinearly on the time variation of

⁴⁶ The tracker field is able to follow the radiation and matter energy densities because the time variation of the tracker-field energy and pressure are controlled by a "frictional effect" that is determined, in turn, by the radiation or matter, (Caldwell, 2000).

the tracker field, which causes novel dynamical features. In the early universe, when the universe is radiation dominated, these features are not evident, and the energy mimics the radiation. However, when the universe undergoes the transition to matter domination, unlike the example discussed above, the k-essence field refuses to track the matter. At first, the field slows down and the kinetic energy density drops sharply, but it soon converges to a fixed value and begins to act as a source of negative pressure. It is then just a short time before it overtakes the matter density (which continues to fall) and drives the universe into cosmic acceleration. The crucial point is that there does not have to be some special feature in the potential energy for this to happen. Rather, it is an automatic dynamical response to the onset of matter domination. See figure 4 below.

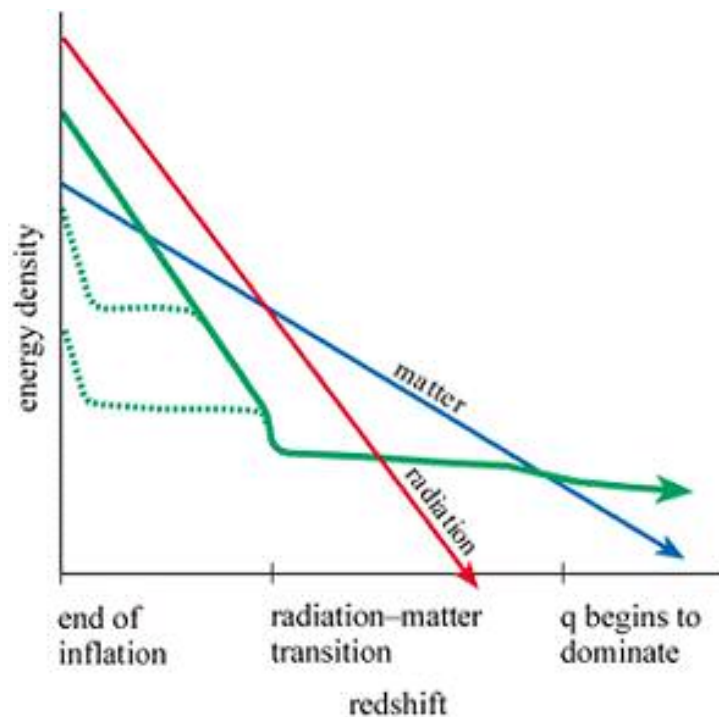


Figure 4: These “tracker models,” in which the amount of quintessence tracks the amount of ordinary matter as the universe expands can explain the missing mass problem and they explain why it appears to be accelerating (see Zlatev, Wang, and Steinhardt (1999)). The diagram is as it appears in Caldwell (2000).

In this picture, the fact that thinking beings and cosmic acceleration occur at nearly the same time in cosmic history is not a coincidence. Both the formation of the stars and planets necessary to support life and the transformation of Quintessence into a negative pressure component are triggered by the onset of matter domination. This is a sufficiently non-anthropic explanation.

After reviewing this, one could say that these tracker models are merely explanations of why we should expect to observe what we observe. In other words, Quintessence is just the right stuff to make us see the right stuff. So how is this superior to the super-string theoretic explanations or any other model from the Identity approach? These Quintessence models are not superior to any Λ models but they do make predictions that can be tested with experiments. In the spirit of the prototypical historical research mentioned earlier, given certain models of Quintessence, and given the nature of the universe and where we're situated in it we ought to be able to see the predicted effects of it. Aside from its ability to explain certain cosmological phenomena, Quintessence, like any good scientific hypothesis, should enable some empirical support if it is to be widely accepted. And it does. With diligent observation we ought to be able to distinguish the predicted effects of Λ from other time-dependent models of dark energy density. Furthermore, observations can be made to distinguish whether Quintessence or a more dynamic model such as k-essence more accurately describes the state of the current universe. The experiments that are underway to discriminate between these hypotheses will be reviewed in detail in the next section. For the time being, let us now turn to the other relevant set of observations that a coherent Eliminativist view must account for, namely the effects of inflation.

For the Identity theorists, even if experimental results would make them concede that another entity (such as Quintessence) may be responsible for current observations, some may still be unwilling to give up the idea that ρ_{vac} is responsible for inflation. If the problem of Λ consisted solely of a numerical inconsistency, the rejection of ρ_{vac} would be unproblematic. However, we don't want to throw the baby out with the bath water. In other words, ρ_{vac} is thought to be good for something; particularly as the driving force behind inflation. And inflation, despite its problematic status, is earning a place in contemporary cosmology. So we should not be so quick to reject ρ_{vac} unless inflation can remain accounted for. For the Eliminativists who reject $\Lambda \equiv \rho_{vac}$, and the cosmological consequences of ρ_{vac} , this point does need to be considered.

Section 6.2: Rethinking Inflation

The Eliminativist inclination to remove ρ_{vac} from the cosmological picture raises the question of how necessary it is for inflationary cosmology. Perhaps a quick overview of inflation will lead us to an answer. If ρ_{vac} proves to be possibly sufficient but ultimately unnecessary to account for inflation, then the Eliminativist approach is a viable response to the cosmological constant problem.

On Inflation

Inflation is shorthand for the idea that in the early universe there was a brief period of rapid exponential expansion. In turn, such a period of extremely rapid expansion may provide a solution to some conceptual problems in standard Big Bang cosmology. When first introduced (1981), Guth's scenario promised to retain many of the

good features of the standard big bang model while overcoming its perceived shortcomings such as: i) the horizon problem, ii) the flatness problem, and iii) the monopole problem.⁴⁷ And as an added bonus the mechanism Guth proposed as the driver of inflation promised to unify cosmology and particle physics.

Original Conception

Originally the idea of inflation was conceived of as a consequence of the vacuum energy associated with a spontaneous symmetry breaking in the Grand Unified Theory (GUT), according to which the strong force and the electroweak force will be unified at energies $\sim 10^{14}$ GeV, corresponding to $\sim 10^{-35}$ seconds after the Big Bang. This was the simplest version of the theory. The huge vacuum energy associated with a GUT symmetry breaking field effectively acts as a cosmological constant that boosts the expansion of the universe exponentially until the vacuum energy is finally converted into heat and the universe enters the standard (slowly expanding) epoch. A quantum-vacuum-driven event with profound cosmological consequences made inflation seem like a clear candidate for a solid connection between cosmology and particle physics. That was the original idea. Yet, Guth's original conception had to be abandoned because it offered no means of a graceful exit and would supposedly leave the universe in too inhomogeneous a state to be consistent with CMBR observations. But given the rapid success and social acceptance of inflation, as well as Guth's own quote that inflation is "too good to be wrong" (1981), newer models were put forth rather than rejecting it.

⁴⁷ As a reminder, these problems concern, respectively, the question of why the background radiation is isotropic, why the observed universe is Euclidean to a very high degree, and why no monopoles are observed (these monopoles are objects that are expected on GUT grounds).

Current Status of Inflation

Much work has gone into finding other versions of inflation that allow for a graceful exit. In order to reconcile these demands with the demand for enough inflation to solve the monopole, horizon, and uniformity problems, inflationary theorists have had to resort to models that appear contrived and/or involve highly speculative elements. To date there are ~ 50 different inflationary models on the market, even more now that Quintessence has been considered as an alternative to ρ_{vac} . Guth optimistically sees these 50 or more models as evidence for the fecundity of the inflation idea; on the other hand it could just as easily be seen as evidence that inflation is not, as originally claimed, an easy and natural remedy for the ills of the standard big bang model. In this respect, Earman and Mosterin claim, “The trouble with ideas that are “too good to be wrong” is that they tend to engender an almost religious faith in their advocates.” They further claim that “Inflationary cosmology has not been able to deliver on its proposed solutions without offering models that are increasingly complicated and contrived, which depart more and more from the standard model it was supposed to improve upon,” (Earman and Mosterin, 2003, pg. 13). These newer models tend to depart from the once attractive features of Guth’s original model that offered a tight connection between cosmology and particle physics by way of GUT symmetry breaking. Inflationary scenarios have been developed showing that such models are in no obvious way related to the spontaneous symmetry breaking of a particle physics model. As early as 1983, Linde suggested a so-called ‘chaotic’ version of inflation that instead of ρ_{gs} , operates with a very simple potential field to drive the inflationary period (Linde, 1990, p.66). The chaotic inflation scenario

thus demonstrated that inflation does not require the kind of quantum field theoretic origin of Guth's original 1981 proposal.

If the cosmological community is determined to keep inflation around for its explanatory potential despite its current conceptual wrinkles, it may be wise to consider such models that do not depend so heavily on vacuum energy as a driving force. Given the unclear status of Λ , if we were to put all of our faith into ρ_{gs} as the source of inflation, then the cosmological constant problem becomes “a potential Achilles heel for the scenario: If some grand principle should be discovered that renders vacuum energy *at all times* impotent, inflation would lose the ultimate source of its power,” (Kolb and Turner, 1993, p. 314).

In later versions of inflation, it is no longer identified with the Higgs field of some GUT. Instead an “inflaton field” is introduced. The ‘inflaton field’ is a fancy way of describing a scalar field to do the work of inflation.⁴⁸ At present the inflaton field is a way to postulate the existence of an entity (or mechanism) that will do the work we want it to do and lead to outcomes consistent with observations but need not be equated with ρ_{vac} per se. Just as quintessence is a scalar field that could be thought of heuristically as “that which produces effects similar to Λ ”, the inflaton field is a scalar field that could be thought of as “that which produces effects similar to ρ_{vac} in the original inflationary conception”.

⁴⁸ The “inflaton field,” as it has come to be called, is an innocent looking scalar field Φ , with a potential $V(\Phi)$. When $V(\Phi)$ is (approximately) constant, the stress-energy tensor reduces to, $T^{\Phi}_{\mu\nu} \approx -g_{\mu\nu}V(\Phi)$, which is the form of a cosmological constant term with $\Lambda_{\Phi} = V(\Phi)/8\pi G$. If Λ_{Φ} is large enough, its inclusion into the Einstein field equations generates the effects of inflation, which in turn produce a spatially flat universe ($k = 0$), (see Earman, 2001)

There are a number of ways of achieving exponential expansion in a cosmological model satisfying Einstein's field equations. For instance, even though it is a relatively old idea, Bludman and Ruderman (1977) argue that even if the vacuum energy density was very large at the time of the symmetry breaking, it was nevertheless negligible in comparison with the thermal energy density of ultra-relativistic particles present at the time. If true, then thermal energy could be inflation's driving force. In 1981, Guth rejected this notion, but has since come to accept it as a possibility, (see Guth, 1997).

More recently, for many cosmologists working on Quintessence, it has lately become a topic of interest as to whether one and the same field could give rise to both inflation and the recently noted dark energy. Such models have been discussed both in the context of standard inflation (Peebles and Vilinkin, 1999) and brane-world inflation (see E.J. Copeland et al., 2001, G. Huey and J. Lidsey, 2001, V. Sahni et al., 2002). As demonstrated by some of these authors, inflation can occur for several of the Quintessence potentials over a well-specified region of parameter space.

Thus far, one cannot say for sure whether these Quintessence-driven inflation models are any better than the original vacuum-driven models. Very few admit that inflation is a wholly unproblematic mechanism, furthermore inflationary cosmologists have never succeeded in producing a model that is free of difficulties. Despite its problems, inflation seems too good to reject by many people standards and is still trying to find its place in standard big bang cosmology. But given that its present epistemic status is still a little unsettled, there may be no profit in demanding that vacuum energy *must* be the source of its power. By adopting alternative models the Eliminativist is able to eliminate ρ_{vac} without eliminating inflation. Hence, the Eliminativist can find a way

around the cosmological constant problem while maintaining the explanatory merits of inflation. Furthermore, these alternative inflationary scenarios would leave different traces of its effect on the cosmos, often depicted as differing signatures in the CMBR. Thus differing inflationary models that are driven by either Λ or by Quintessence can also be tested against each other in an experimental situation. Experimentation could undoubtedly resolve some of the present issues regarding the cosmological significance of vacuum energy, and by knowing the extent of its cosmological significance, if any, we would be in a better position to resolve the cosmological constant problem.

Section 7: The Experimental Approach to the Cosmological Constant Problem

Testing Competing Models in Experimental Cosmology

As has been stated in previous sections, the cosmological constant problem springs from a numerical mismatch of 120 orders of magnitude between the typical energy scales theorized by particle physicists on one hand and the observational constraints imposed by cosmologists on the other. Various schemes have been devised to provide a way of canceling out the vacuum energy. The motivation behind the development of such schemes comes from the presupposition that when we speak of ρ_{gs} and when we speak of Λ , we are speaking of one and the same thing that has mysteriously decreased. When we consider the key features of the current universe (its mass content, acceleration, etc.) all attempts to attribute them to Λ lead to the further claim that $\Lambda \equiv \rho_{gs}$, creating the cosmological constant problem. Surely this whole problem could dissolve in an instant if we ever realized that something other than vacuum energy is responsible for these observations. We cannot say for certain that “ Λ ” qua vacuum energy has been observed, but rather *something* – some kind of dark and unknown energy – has been observed that accounts for roughly two thirds of the universe’s contents and is making it accelerate.

In light of the arguments made in the previous sections, perhaps some reevaluation of the problem as a problem is needed. As we have seen, by identifying Λ with ρ_{gs} , the Identity approach requires some cancellation scheme or something to that effect to explain why ρ_{gs} is so small. For the Identity theorists, to solve the cosmological constant problem would be to do something like this. The Eliminativist approach on the other hand poses an alternative to Λ by turning to other models of dark energy such as

Quintessence to account for inflation and the recent observations, whereby the whole inconsistency is circumvented. These are both viable approaches and at this stage in scientific cosmology, the debate between these approaches can be settled by experiment. For the time being, instead of treating the cosmological constant problem as one that requires a variety of theoretical cancellation or elimination maneuvers in order to make progress, let us treat the cosmological constant problem experimentally. Why not turn to experimentation for progress with the problem? After all, as we saw in section 2, cosmology has the means to make observations that can support counterfactual statements relating to theories. By allowing for experimental situations cosmology can bring relevant observations to bear on competing hypotheses to discriminate between them. Through diligent experimentation we can determine whether Quintessence or Λ models more accurately describe our universe, and in so doing, we become clearer about which direction an adequate solution to the cosmological constant problem lies.

At present, due to the relative novelty of Quintessence models, not enough evidence has accumulated to rule in favor of it over Λ . The results of Perlmutter, Turner, and White (1999) indicate a favorable outcome for Λ over quintessence. However, Huey et al. (1999) conclude that new observational techniques are necessary to rule out Quintessence in favor of Λ . There are currently various techniques underway that could either support Quintessence, or at least determine how it may stand up against Λ . Here I will only state a few.

If we interpret Quintessence and Λ as fluids, they have different equations of state and different values of the equation of state parameter, w (recall equation 55). More precise measurements of supernovae over longer distances may be able to separate these

two possibilities. Much work has been done to this end. Recalling section 4.3, the supernova observations of Perlmutter et al. (1999) fit best with a model that suggest Λ or something like it is non-zero. See figure 5 below.

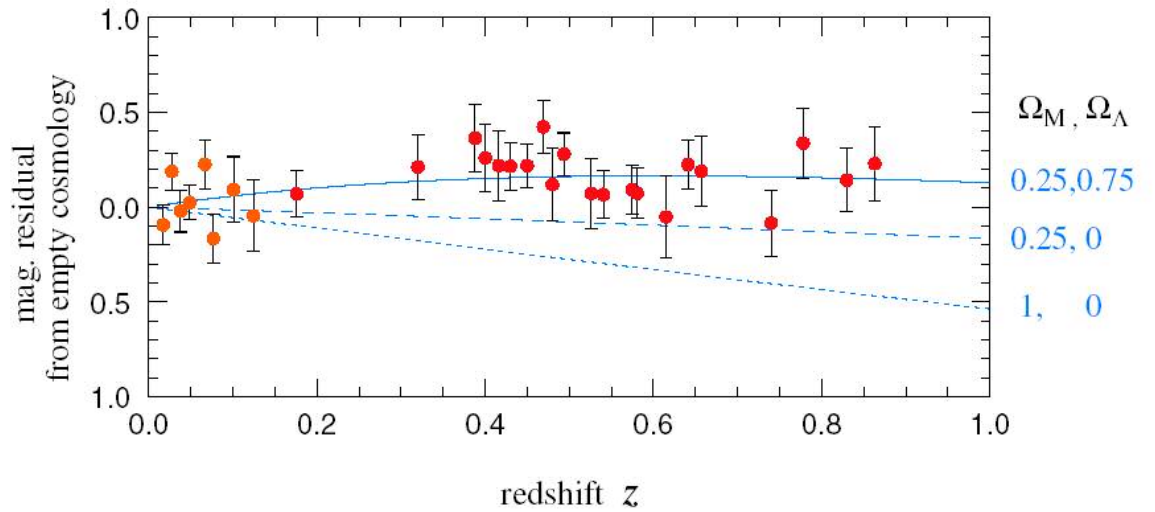


Figure 5: From the same collaboration (the Supernova Cosmology Project, Knop et al. (2003)), we see that the data indicate a multitude of supernova samples that are in agreement with a model where $\Lambda > 0$.

With these data, Λ and quintessence models have been constructed and with more supernova data of this kind at higher redshifts we will be able to differentiate between them by their equations of state, (see Saini et al., 2000). In their paper, “*Reconstructing the Cosmic Equations of State from Supernova Distances*” (2000), these authors recognize that the original Λ term could be described as an effective scalar field with some potential $V(\phi)$ that is minimally coupled to the gravitational field and has little or no coupling to other known physical fields. For them, since the more fundamental theories like supergravity or M theory can provide a number of possible candidates for the ‘ Λ -field’ but do not uniquely predict its potential $V(\phi)$, they decided to reconstruct it from present-day cosmological observations. From this potential of the field we can

obtain its equation of state (recall equation 56). If the potential is assumed to be dynamic with different values at different times and redshifts, then its equation of state at different times and redshifts can be determined as well. Their reconstruction for $w_\phi(z)$, i.e. the equation of state as a function of redshift, is plotted in figure 6 below.

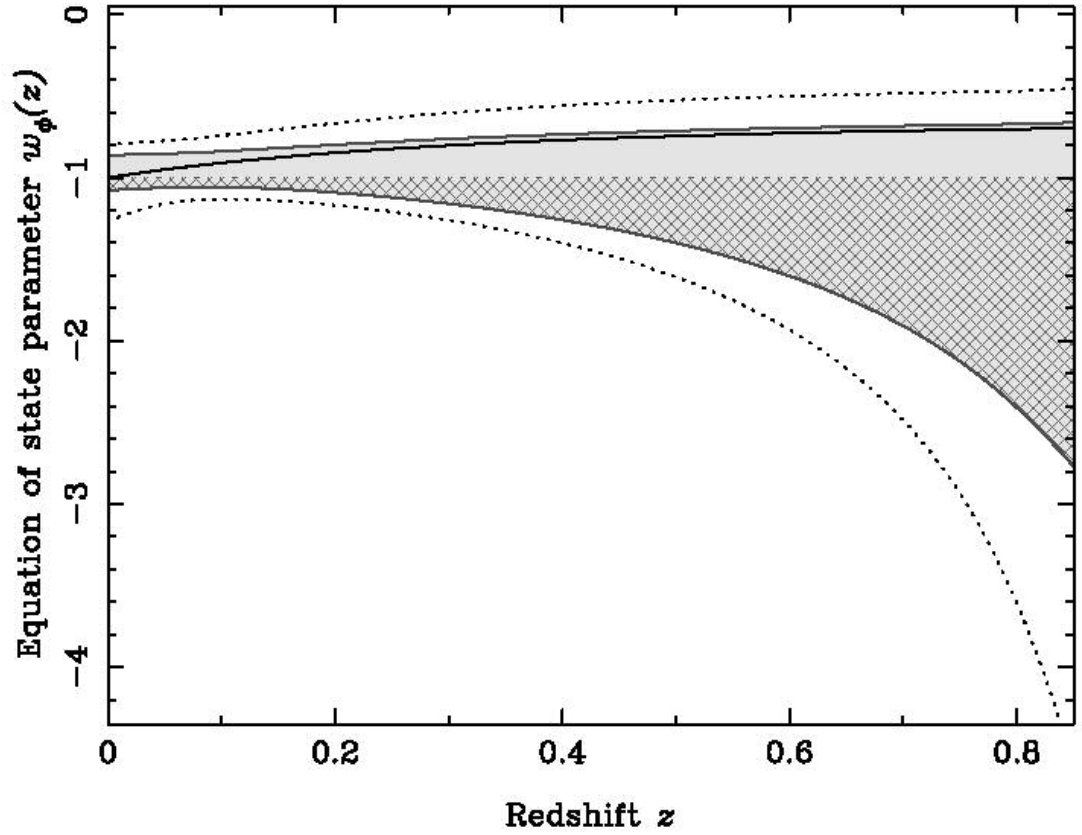


Figure 6: The equation of state of dark energy/quintessence is reconstructed from observations of Type Ia high redshift supernovae in a model independent manner. The equation of state parameter $w_\phi(z)=P/\rho$ is seen as a function of redshift. Assuming $\Omega^M \approx 0.3$, the solid black line corresponds to the best-fit values of the parameters. The shaded area covers the range of 68% errors, and the area between the dotted lines covers the range of 90% errors. There is some evidence of possible evolution in w_ϕ with $-1 \leq w_\phi \leq -0.86$ preferred at the present epoch, and $-1 \leq w_\phi \leq -0.66$ at $z = 0.83$, the farthest supernova in the sample. This model describes a universe whose dark energy content has a dynamic equation of state. However, a cosmological *constant* with $w = -1$ is consistent with the data as well, (from Saini et al., 2000).

From their work we now have a model that relates *supernova observations* as a function of redshift with *equations of state* as a function of redshift. This relation is key for distinguishing between Λ and quintessence models since we can distinguish these dark energy candidates by their equations of state. Sahni et al. (2003), have developed a technique for doing this by using what they call the Statefinder statistic⁴⁹. According to Sahni “An issue of the utmost importance is whether dark energy (equivalently quintessence) is a cosmological constant or whether it has a fundamentally different origin. A new dimensionless statistic ‘Statefinder’, recently introduced ... has the power to discriminate between different forms of dark energy and may therefore be a good diagnostic of cosmological models,” (2003, pg. 9). Essentially, the Statefinder statistic (actually a pair of numbers) allows us to characterize the properties of dark energy in a model independent manner. The Statefinder pair is calculated for a number of existing models of dark energy having both constant and variable equations of state. See the figure 7 below.

⁴⁹ Introduced by Sahni et al. (2003), the Statefinder statistic is a new cosmological diagnostic pair $\{r, s\}$. The Statefinder is a geometrical diagnostic that allows them to characterize the properties of dark energy in a model independent manner. The parameter r forms the next step in the hierarchy of geometrical cosmological parameters after the Hubble parameter $H = da/a$, and the deceleration parameter $q = -ad^2a/da^2$ and is represented as $r = d^3a/aH^3$, while s is a linear combination of q and r chosen in such a way that it does not depend upon the dark energy density and is represented as $s = (r-1)/(3(q-1/2))$. The Statefinder pair $\{r, s\}$ is algebraically related to the equation of state of dark energy and its first time derivative.

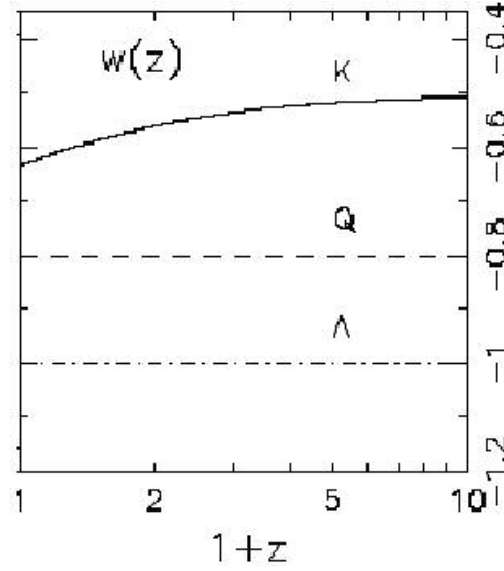


Figure 7: Sahni et al. (2003) have shown the different equations of state as a function of redshift for K-essence, Quintessence, and Λ models, respectively.

The Statefinder statistics used to model these equations of state can be determined to very high accuracy from a SNAP-type experiment, (Sahni et al., 2003). SNAP is the space-based *Supernova Acceleration Project* that will monitor the sky for supernovae and other time-varying astrophysical phenomena. We can hypothesize in a counterfactual way that if the equation of state of this dark energy were dynamic and were within a specified range of values at such and such a redshift, we would expect to see a well specified distribution of supernovae at that red shift. Once this experiment gets underway and can probe the universe for type Ia supernovae at higher redshifts, there will be plenty of data to support one of these equations of state over the others, and will give us an indicator of the kind of dark energy that is responsible for the current acceleration and missing mass.

Another technique under consideration for distinguishing models of Quintessence from Λ is by appealing to the effects of acceleration. Cosmic acceleration would affect

the number of galaxies we would expect to find as we look deeper and deeper into space. With appropriate corrections for evolution and other effects, the average density of galaxies is uniform throughout space. Consequently, for a fixed range of distances, one should find the same number of galaxies nearby and far away. But cosmologists measure the redshift of distant galaxies, not their distance. The conversion from redshift to distance follows a simple linear relation (the Hubble law) if the distances are small, but a nonlinear relation depending on the acceleration of the universe if the distances are large. The nonlinear relation will cause the number of galaxies found for a fixed range of redshifts to change systematically as one probes deeper into space. Once again, depending on the nature of the dark energy that is causing the universe to accelerate, a Quintessence model would predict a slightly different non-linear relation than a Λ model. If Quintessence really were the source of the current acceleration we would expect to see a different relation between the number of galaxies and redshifts than the relation expected if Λ were the source of the acceleration. The Deep Extragalactic Evolutionary Probe (DEEP), an advanced spectrograph on the Keck II telescope in Hawaii, is poised to test this prediction with an accuracy that may be sufficient to distinguish between Quintessence and a cosmological constant, (see Caldwell, 2000, pg. 9).

Finally, cosmic microwave background experiments that are planned or in progress will also be able to distinguish between Λ models and a wide range of Quintessence models. Λ and Quintessence make different predictions about the observable state of the CMBR. It is believed that Quintessence can gravitationally clump at very large scales, while vacuum energy density, as we're currently prone to think of it, would not display this feature. If so, Quintessence could gravitationally interact with

ordinary matter under certain conditions (and should have already). Ordinary particles are physically very small compared with the Compton wavelength of the Quintessence particles, so individual particles have a completely negligible effect on Quintessence (and vice versa). However, very large clumps of ordinary matter, spread out over a region comparable with the Compton wavelength, can interact gravitationally with Quintessence and create inhomogeneities in its distribution that may produce detectable signals in the cosmic microwave background. Small variations in the amount of Quintessence across the sky should be seen as ripples in the CMBR temperature that should, in principle, be observable and distinguishable from ripples allegedly created by quantum vacuum fluctuations, (see Kawasaki, Moroi, and Takahashi, 2001). Fortunately for experimental cosmologists, CMBR is all around us, and the interactions (or interventions) between Quintessence and ordinary matter, if existent, would be detectable in a multitude of different instances.

In these experimental examples there are plenty of data to be obtained. As mentioned earlier, counterfactual support in experimental situations need not require literal interventions if nature is generous and provides us with enough data from the system being observed. In these cases, whether we are directing our observations toward localized regions of the CMBR, supernovae or galaxy clusters, we can test these hypotheses by considering what we would expect to see if the conditions under which we make our observations are varied. Fortunately for us the cosmos has a lot of these things, where variations are likely to exist and can be observed. In time such experiments may rule out Quintessence as a viable alternative to Λ or vice versa. As experimentation has helped cosmology in the past to progress this far, it gives us hope of resolving current

issues such as the cosmological constant problem. With these developments in experimental cosmology during the recent years, we are now in a position to get an even more precise understanding of the cosmic energy that is responsible for what we observe. If these experiments rule out the cosmological constant as a dark energy candidate, then that will essentially dissolve of the problem. In effect, to eliminate the cosmological constant is to eliminate the cosmological constant problem. If experimentation leads to evidence in favor of vacuum energy as the most probable dark energy candidate, then some physicists, such as the Identity theorists, will be much clearer about the work that lies before them for telling a consistent story of Λ .

Section 8: Summary and Conclusion

The history of the cosmological constant seems to be shadier than any other constant in physics. It has been both debased and praised, and has been ruled out many times, only to resurface time and again. Its varying status has been accompanied by varying attitudes. At one extreme there are those who, like Einstein, have felt that a cosmological term in the gravitational field equations is an ugly addition that could not be realized in nature. At the other extreme is the view that, whether we like it or not, Λ as vacuum energy density is an inevitable feature of nature.

Between these two ends of the spectrum is the more pragmatic stance whereby Λ is brought in only when it is needed to fix some relevant cosmological problem. Much of Λ 's career has been spent as a cosmological fixer: first used by Einstein to make possible a static general relativistic model of the universe, it was later pulled off the shelf again to solve the age problem, to explain structure formation, and then as the explanatory mechanism for the redshifts of quasars. Then with the rise of inflationary cosmology it was invoked to overcome the horizon problem and other explanatory inadequacies of standard big bang cosmology. Our present conception of the cosmos requires a positive Λ , or something like it, to explain the accelerating expansion and to account for the missing mass required to ensure spatial flatness. For many cosmologists, Λ may be here to stay for a while, if so, the cosmological constant problem will need to be addressed.

Historically, the cosmological constant problem has gone through various stages as a problem. The progression of the changes began when Zel'dovich suspected in the late 1960s that vacuum energy may have gravitational effects that we would be remiss to ignore. This led him and others to interpret ρ_{vac} as a non-zero cosmological constant. In

an attempt to derive Λ from physical constants, a discrepancy emerged with observational constraints. Later, in the early 1970s as theoretical physicists developed their hypotheses about the early universe that suggested electroweak unification and spontaneous symmetry breaking the discrepancy grew to ~ 120 orders of magnitude. Then with the advent of inflationary cosmology in the early 1980s, a large vacuum energy density was no longer a perplexing notion to be explained, but became a required feature to provide the cosmos with the initial boost to instigate inflationary expansion. The problem went from explaining a numerical inconsistency to explaining how such a large numerical value mysteriously vanished after a specific phase in the early universe. In the 1990s new observations indicating that roughly 70% of the cosmos is filled with an unknown or ‘dark’ energy that is causing it to accelerate led some cosmologists to designate Λ as the dark energy candidate accountable for these observations. This *explicitly* suggests $\Lambda \neq 0$ and so whatever cancellation scheme is devised must be precise to 120 decimal places.

To some the cosmological constant problem is an epistemic problem in that there must be some mechanism that is responsible for the vanishing of Λ , and we just have not figured out all the details yet. The cosmological constant problem could also be characterized as an ontological problem in that the difficulty of explaining how Λ has vanished stems from the difficulty in reconciling the ontological status of Λ in QFT with its status in GR, and that no solution can be properly posed until we have a better sense of the ‘cosmic stuff’ that Λ allegedly represents. For almost four decades, the cosmological constant problem has been a thorn in the side of modern physics, and has created a virtual

industry of solution types. The basic solution types that were addressed here are the Identity Approach and the Eliminativist approach. The Identity Approach claims that $\Lambda \equiv \rho_{gs}$, and so quantum ground state energy is responsible for all of these observed effects. The Eliminativist approach is not committed to such a claim. In fact the Eliminativist denies the reality of ρ_{gs} and therefore denies all phenomena ρ_{gs} is responsible for, or seeks to provide a different explanation. This distinction in starting positions leads to very different solution strategies. Regardless of one's intuitions regarding the ontological status of Λ , any approach to solving the problem must be able to give an explanation for the inconsistency while at the same time accounting for all the observations.

Since the inception of the problem in the late 1960s until the observations made in the late 1990s much of the work done to solve the problem was basically Identity-driven. It meant trying to find some cancellation mechanism to explain how all the vacuum energy disappeared, and if such a mechanism could be adequately described, the problem would be solved. Solutions were sought through supersymmetry, super gravity, then eventually string theory. Even though an Identity type solution can be sought by treating the problem semiclassically, efforts to dissolve the problem were also sought this way. Some of these attempts were more successful than others, but they were and still are, at best, promissory notes. The problem until then was a problem of resolving an inconsistency; vacuum energy was not thought to have any *current* cosmological significance. However, once the recent observations were made, these Identity theorists designated vacuum energy as the unknown dark energy responsible for all of these observed effects. The Eliminativist explicitly rejects this. While the Eliminativists need

not doubt the recent observations, they attribute another cause to them; instead of the cosmological constant, the observations may be attributable to another dark energy candidate such as Quintessence. If something else is responsible for the universe's early inflationary period as well as its 70% missing mass and acceleration, then the cosmological constant could indeed be zero, and accounting for the 120-order inconsistency could be disregarded as a meaningless problem. The Eliminativist rejects the cosmological constant and in so doing eliminates the cosmological constant *problem*.

The fact that these two approaches, Identity and Eliminativist, have emerged from the recent observations is significant because it now allows physicists to take an experimental approach to the cosmological constant problem. At this stage in cosmology's development as a science we have seen that it need not be thought of as merely an observational science, but that it is capable of testing the hypotheses of its domain with natural experiments as effectively as any other science. This ability to conduct experiments is especially useful given the current status of the problem. When we ask the questions of 'what is the universe made of?' and 'what's causing it to accelerate?', these can be answered experimentally. Both the Identity theorists and the Eliminativists have different answers as to what the energy source is that is responsible; their dark energy candidates can be tested against one another experimentally.

Technologies are currently underway to probe deeper into the cosmos and make more precise measurements of supernovae and of the CMBR. We are fortunate to live in a universe that enables us to gather large amounts of data that can bear upon the competing hypotheses regarding its contents and dynamics. The outcome of such experimentation would be greatly beneficial to anyone concerned about the cosmological

constant problem. If experimentation can be done to rule out vacuum energy as a dark energy candidate, then we may send Λ back to the shelf once again and declare the cosmological constant problem dissolved. If experimentation leads to evidence in favor of vacuum energy as the most probable dark energy candidate, then at least this will limit the scope of the problem and physicists will be much clearer about the work ahead of them in order to give a consistent account of Λ .

Appendix

The Anthropic Approach to Solving the Cosmological Constant Problem

Anthropic explanations fall somewhat outside of the kinds of explanations we would consider as satisfactorily scientific, but since they seem to attract considerable interest in this context of the cosmological constant problem, they deserve some consideration. Anthropic arguments are easily misused, and have often been employed as a way to get around the difficult task of understanding the real reasons behind why we observe the universe we do. They often amount to, at best, an account that is purely descriptive but ultimately vacuous. To say that the reasons why we observe the value of Λ that we do are because if it were different we would not be here seems strange and not very explanatory in the conventionally scientific sense – how can the fact that the universe is hospitable to observers constitute an explanation of anything? We will consider two different anthropic models that are intended to solve the cosmological constant problem by Weinberg (1997) and Vilenkin (2001). And as will be shown, aside from stretching the limits of what is deemed acceptable as a scientific explanation, neither argument seems to fully address all the critical aspects of the problem at hand. Also, the anthropic approach does not distinguish between the Identity and the Eliminativist views and so the power of experiment to gather evidence in favor or against one of these views does not relate here.

On the Anthropic Principle

The Anthropic Principle was first presented in 1973, by the astrophysicist and cosmologist Brandon Carter, at a conference held in Poland to celebrate the 500th birthday of Nicolaus Copernicus. The Anthropic Principle is an attempt to explain the observed fact that the fundamental constants of physics and chemistry are just right or finely tuned to allow the universe and life as we know it to exist. The Anthropic Principle says that the seemingly arbitrary and unrelated constants in physics have one strange thing in common – these are precisely the values you need if you want to have a universe capable of producing and sustaining life. Everything from the particular energy state of the electron to the exact level of the weak nuclear force seems to be tailored for us to exist. We appear to live in a universe dependent on several independent variables where the slightest change would render it inhospitable for life to develop. And yet, here we are. The anthropic principle states that the reason we are here to ponder this question at all, is due to the fact that *all* the correct variables are in place. Later, in 1983, Carter claimed that, in its original form, the principle was meant as a caveat for astrophysicists and cosmologists to be mindful of possible errors in the interpretation of astronomical and cosmological data unless the biological constraints of the observer were taken into account. Yet, as a principle of science, its reception has been lukewarm at best.

Carroll notes, “Many professional cosmologists view this principle in much the same way as many traditional literary critics view deconstruction as somehow simultaneously empty of content and capable of working great evil,” (2000, pg. 30). And Weinberg, despite his zeal toward it, has remarked that a physicist talking about the anthropic principle runs the same risk as a cleric talking about pornography: no matter

how much you say you're against it, some people will think you're just a little too interested.

The anthropic principle suggests that there may be parameters characterizing the universe we observe that may not be determined directly by the fundamental laws of physics. It is based on the truism that intelligent observers will only ever experience conditions which allow for the existence of intelligent observers, (for further reading, see Barrow and Tipler (1986) and Hogan (1999)). But according to critics, this is simply a tautology, a very elaborate way of saying "if things were different, they would be different". In order for the tautology that 'observers will only observe conditions which allow for observers' to have any force, it is necessary for there to be alternative conditions or alternative parts of the universe where things are different. In such a case, our local conditions arise as one out of many possible combinations from a relative abundance of different environments and the likelihood that such environments would give rise to intelligence. In 1983, Linde's model of chaotic inflation was based on conditions of just this sort. It was by appealing to the anthropic principle that he was able to explain why the initial conditions of our universe were such that inflation was possible.

The proposal of an anthropic solution to the cosmological constant problem mostly concerns the idea that our universe is embedded in a larger structure (a 'multiverse'), and that we live in a particular universe in which the cosmological constant is compatible with conditions for life forms to evolve (Weinberg, 1989, 1997). This ensemble, or multiverse, may consist of different expanding regions at different times and locations in the same multidimensional spacetime. If the vacuum energy density varies among the different members of this ensemble, then the value observed by any species of

astronomers will be conditioned by the necessity that this value of ρ_{vac} should be suitable for the evolution of intelligent life. Weinberg has further remarked “It would be a disappointment if this were the solution of the cosmological constant problems, because we would like to be able to calculate all the constants of nature from first principles, but it may be a disappointment that we will have to live with,” (2000, pg. 4). While Weinberg has continually made use of anthropic reasoning in connection with the cosmological constant problem, and presently sees the anthropic line of reasoning to the cosmological constant problem as the most promising (see Weinberg, 1997 and Cao, 1999), other physicists are much less convinced by this way of thinking. Nevertheless the anthropic solution types are presented as follows.

Anthropic Solutions

Weinberg's view

The anthropic bound on a positive vacuum energy density is set by the requirement that ρ_{vac} cannot be so large as to prevent the formation of galaxies. However, we would not expect to live in a big bang in which galaxy formation is just barely possible. A much more reasonable idea is characterized by what Vilenkin calls the principle of mediocrity, (see Vilenkin, 1995). This principle suggests that we should expect to find ourselves in a big bang universe that is typical among those where the development of intelligent life is both possible and probable. To be specific, if $P(\rho_{vac})d\rho_{vac}$ is the a priori probability of a particular big bang having vacuum energy density in the range between ρ_{vac} and $\rho_{vac} + d\rho_{vac}$, and $N(\rho_{vac})$ is the average number of

scientific civilizations in big bang universes with energy density ρ_{vac} , then the probability of a scientific civilization observing an energy density between ρ_{vac} and $\rho_{vac} + d\rho_{vac}$ is

$$dP(\rho_{vac}) = N(\rho_{vac})P(\rho_{vac}) d\rho_{vac} \quad (I)$$

Summarizing over the technical parts of Weinberg's argument, he concludes that the probability of finding ourselves in a big bang with a vacuum energy density large enough to give a present value of ρ_{vac} (or Λ) of ~ 0.7 turns out to be $\sim 5\% - 12\%$, depending on the assumptions used. "In other words, the vacuum energy in our big bang still seems a little low, but not implausibly so. For Weinberg, these anthropic considerations can provide a solution to the cosmological constant problem, provided of course that the underlying assumptions are valid. But then again, setting the correct initial conditions to place us in a region that's friendly to intelligent observers does not require any more fine-tuning than other non-anthropropic solutions that arise in say, string theory.

As much as one can appreciate the precision of Weinberg's mathematical approach to an anthropic solution, an immediate issue come to mind. First of all, the use of multi-verse scenarios could make some people skeptical and cautious about anthropic reasoning. If a solution to the cosmological constant problem were devised by appealing to anthropic reasoning in a multi-verse scenario, employed by certain models of inflation, this would undermine the usual observational basis for a scientific explanation. But this feature is no different than brane-theoretic solutions. Hypotheses regarding the states of different universes do not have a chance at being empirically tested or being empirically adequate, to say the least. Regarding the ensemble of other universes in the multi-verse, one is *in principle* unable to verify their existence observationally since we cannot, by definition, be in causal contact with these other universes as they lie outside our horizon,

(that is, outside our light-cone). One might argue that if a multi-verse model is supported by means of observations in our universe, then it is not unfair to speculate on conditions in other universes. However, due to the unobservability of these hypothetical ‘other’ universes, this effectively amounts to making statistical arguments based on only one data point, i.e. the conditions in *our* Universe, (Rugh and Zinkernagel, 2002).

Weinberg is not alone on this anthropic crusade to solve the cosmological constant problem. Another anthropic argument has been put forth by Alexander Vilenkin.

Vilenkin's View

In his paper “*Cosmological constant problems and their solutions*” (2001), (which was written with a notable Weinberg influence) Vilenkin says, “I realize that the anthropic approach has a low approval rating among physicists. But I think its bad reputation is largely undeserved,” (pg. 1). His approach is different from Weinberg’s regarding how much he postulates. His approach is anthropic but does not postulate the existence of an ensemble of other universes in a multiverse. For Vilenkin, what we perceive as the cosmological constant is in fact a stochastic variable that varies on a very large scale, greater than the present horizon, and takes different values in different parts of the universe, and that a situation of this sort can naturally arise in the context of a well defined inflationary scenario.

The key observation here is that gravitational clustering has led to galaxy formation. An anthropic bound on ρ_{Vac} can be obtained by requiring that it does not dominate before the redshift z_{max} when the earliest galaxies are formed. With $z_{max} \sim 5$ one obtains

$$\rho_{vac} < 200\rho_{M0} \quad (II)$$

where ρ_{M0} is the present matter density. If all values in the anthropic range (equation II) were equally probable, then, as Vilenkin claims, $\rho_{vac} \sim \rho_{M0}$ would still be ruled out at a 95% confidence level. However, the values in this range are not equally probable. The anthropic bound (equation II) specifies the value of ρ_{vac} that makes galaxy formation barely possible. Most of the galaxies will not be in regions characterized by these marginal values, but rather in regions where ρ_{vac} dominates after the bulk of galaxy formation has occurred. Thinking about this more quantitatively, a probability distribution could be represented as

$$dP(\rho_{vac}) = P^*(\rho_{vac})v(\rho_{vac})d\rho_{vac} \quad (III)$$

Here, $P^*(\rho_{vac})d\rho_{vac}$ is the prior distribution that is proportional to the volume of those parts of the universe where ρ_{vac} takes values in the interval, $d\rho_{vac}$ (or rather, ρ_{vac} and $\rho_{vac} + d\rho_{vac}$) and $v(\rho_{vac})$ is the average number of galaxies that form per unit volume. The distribution of equation (III) gives the probability that a randomly selected galaxy is located in a region where the cosmological constant is in the anthropically friendly interval ρ_{vac} and $\rho_{vac} + d\rho_{vac}$. If we are typical observers in a typical galaxy, then we should expect to observe a value of ρ_{vac} somewhere near the peak of this distribution. The distribution $P^*(\rho_{vac})$ is determined from the inflationary model of the early universe. Thus, according to Vilenkin, all one needs is a particle physics model that would allow ρ_{vac} to take different values and an inflationary model that would give a more or less flat prior distribution $P^*(\rho_{vac})$ in the anthropic range, (2001, pg. 7). In this case, we exist as observers in a galaxy whereby we are quite likely to observe the value of Λ that we do.

In this anthropic approach, instead of postulating claims about how likely our universe is to have the features it has in comparison with others in the ensemble, Vilenkin speaks more locally about how likely our galactic neighborhood is to have the features it has compared to other regions in our universe. This model of the universe is, of course, a deviation from the once thought homogeneous distribution of matter and energy.

Vilenkin's universe is much more segregated. Instead of a smooth cosmic ball, perhaps a more fitting analogy would be like that of a soccer ball, with an evenly distributed amount of black pentagons on a white sphere, where the black pentagons represent the dense regions. Our galaxy could then be thought of as a mere speck on one of those black pentagons, and even though everything in our surroundings seems dense, this density does not spread out evenly across the universe. This density distribution is much more localized than we once thought.

Vilenkin's approach is an interesting alternative to Weinberg's multi-verse scenario, but nevertheless falls subject to the same issue of empirical testability and support. Even though the energy density is localized, the stochastic variations that ρ_{vac} would allow in this scenario only occur, as Vilenkin states, "on a very large scale, greater than the present horizon" (pg. 6). This means that we are not in causal contact with these other regions. We could not verify, even in principle, whether just over our cosmic horizon ρ_{vac} decreases and then increases in yet a more distant region.

In conclusion, anthropic reasoning seems to radically change what it means to give an effective explanation within the physical sciences. Understandably, a sense of disappointment would inevitably accompany the realization that there were limits to our ability to unambiguously and directly explain the observed universe from first principles.

Regardless of the controversial status of anthropic reasoning for providing a scientific explanation of the cosmological constant, the very fact that it has been frequently invoked as a possible solution type to the problem is an indicator of how serious and perplexing the cosmological constant problem really is.

Bibliography

- 1) Armendariz, C., Mukhanov, V., and Steinhardt, P. J. (2000) xxx.lanl.gov/abs/astro-ph/0004134
- 2) Banks, T. (1999), "*M theory and cosmology*," hep-th/9911067
- 3) Banks, T. (1996) "*SUSY Breaking, Cosmology, Vacuum Selection and the Cosmological Constant in String Theory*," hep-th/9601151.
- 4) Barrow, J.D., and Tipler, F.J. (1986), The Anthropic Cosmological Principle (Oxford: Clarendon,).
- 5) Bernard C., (1865) Introduction to the Study of Experimental Medicine
- 6) Birrell, N.D. and Davies, P.C.W. (1982) Quantum Fields in Curved Space (Cambridge: Cambridge University Press)
- 7) Bludman, S.A and Ruderman, M.A. (1977) "*Induced Cosmological Constant Expected above the Phase Transition Restoring the Broken Symmetry*" Physical Review Letters 38, 255-257
- 8) Bohr, N. (1948) "*Problems of elementary-particle physics*" (Reports of an International Conference on Fundamental Particles, Cavendish Laboratory, Cambridge July 1946). (Reproduced in [31], p.219-222)
- 9) Bohr, N. (1949) '*Discussion with Einstein on Epistemological Problems in Atomic Physics*' in The Philosophical Writings of Niels Bohr – Vol.II, Essays 1932-1957 on Atomic physics and Human Knowledge, Reprint 1987, (Connecticut: Ox Bow Press) [originally, Wiley 1958], pp. 32-66
- 10) Bohr, N. and Rosenfeld, L. (1933) '*Zur Frage der Messbarkeit der Elektromagnetischen Feldgrößen*', Mat.-fys.Medd.Dan.Vid.Selsk. 12, no. 8. This paper appear in translation in J.A. Wheeler and W.H. Zurek (eds), Quantum Theory and Measurement (Princeton, NJ: Princeton University Press, 1983), pp. 465 - 522.
- 11) Brax, P.H., (2004) "*The Cosmological Constant Problem*", Contemporary Physics, 45, 3, 227, 2004.
- 12) Brecher, K. and Silk, J. (1969) "*Lemaître Universe, Galaxy Formation and Observations*," Astrophysical Journal **158**: 91–102.
- 13) Caldwell, R., (2000) "*Quintessence*", PhysicsWeb – Physics World,

- 14) Calzetta, E., and Hu, B. (2005) *Noise and Fluctuations in Semiclassical Gravity*, arXiv:gr-qc/9312036
- 15) Cao, T.Y. (1999) (ed) Conceptual foundations of quantum field theory (Cambridge: Cambridge University Press) Contributions to conference on '*Historical and Philosophical Reflections on Quantum Field Theory*' held at Boston University, March 1-3, 1996.
- 16) Carroll, S.M. (1998) "*Quintessence and the Rest of the World*," Physical Review Letters, **81**: 3067–3070.
- 17) Carroll, S.M. (2000) "*The Cosmological Constant*", astro-ph/0004075.
- 18) Cleland, C.E., (2002) "*Methodological and Epistemic Difference between Historical Science and Experimental Science*," Philosophy of Science, 69, 474
- 19) Collingwood, R.(1940): *An Essay on Metaphysics*. Oxford: Clarendon Press.
- 20) Copeland, E. J., Liddle, A. R. and Wands, (1998) D., Phys. Rev. **D57**, 4686 grqc/9711068.
- 21) Copeland, E.J., Liddle, A.R, and Lidsey, J.E. ,(2001) Phys. Rev. D 64 023509.
- 22) Cornish, N.J. and Starkman, G. (1998) "*Of Matter Less Repulsive than a Cosmological Constant*," astro-ph/9805027.
- 23) De Sitter, W. (1917) "On the Curvature of Space," *Proceedings of the Section of Sciences, Koninklijke Akademie van Wetenschappen te Amsterdam* **20**: 229–242.
- 24) Dirac, P. (1938), Proceedings from the Royal Society, 165, 199.
- 25) Dreitlein, J. (1975) '*Broken Symmetry and the Cosmological Constant*' Physical Review Letters 33, 1243-1244
- 26) Earman, J. (2001) "*Lambda, The Constant that Refuses to Die*," Arch. Hist. Exact Science, 55, 189-220.
- 27) Earman, J. and Mosterin, J. (1999) '*A Critical look at Inflationary Cosmology*' Philosophy of Science, 66, 1-49
- 28) Earman, J., (2003) '*The cosmological constant, the fate of the universe, unimodular gravity, and all that*', Studies in History and Philosophy of Modern Physics 34, 559
- 29) Eddington, A.S. (1930) "*On the Instability of Einstein's Spherical World*," Monthly Notices of the Royal Astronomical Society, **90**: 668–678.

- 30) Eddington, A.S. (1931), Proceedings from the Royal Society, 133, 605.
- 31) Eddington, A.S. (1933) The Expanding Universe, New York: Macmillan.
- 32) Einstein, A. (1917a) "*Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie*," Königlich Preussische Akademie der Wissenschaften (Berlin). *Sitzungsberichte*. 142–152. [English translation as "Cosmological Considerations on the General Theory of Relativity," in Perrett and Jeffrey (eds.), *The Principle of Relativity*, 177–188. New York: Dover Books, 1952.]
- 33) Einstein, A. (1917b) "*Über die spezielle und die allgemeine Relativitätstheorie*." Braunschweig: View eg. English translation as *Relativity: The Special and the General Theory*. New York: Crown Publishers.
- 34) Einstein, A. (1922) "Bemerkung zu der Arbeit von A. Friedmann: "Über die Krümmung des Raumes," *Zeitschrift für Physik* **11**: 326.
- 35) Einstein, A. (1931) "*Zum kosmologischen Problem der allgemeinen Relativitätstheorie*," Preussische Akademie der Wissenschaften (Berlin). *Sitzungsberichte*. 235–237.
- 36) Enz, C.P. and Thellung, A. (1960) '*Nullpunktsenergie und Anordnung nicht vertauschbarer Faktoren im Hamiltonoperator*', *Helv. Phys. Acta* **33**, 839-848
- 37) Ferreira, P. G. and Joyce, M., (1998), *Phys. Rev.* **D58**, 023503, astro-ph/9711102
- 38) Friedmann, A. (1922) "*Über die Krümmung des Raumes*," *Zeitschrift für Physik*, 10: 377–386.
- 39) Gamow, G. (1970) My World Line, Viking Press, NY.
- 40) Gasking, D. (1955): "Causation and Recipes", *Mind*, **64**, pp. 479-487.
- 41) Green, M. B., Schwarz, J. H., and Witten, E., (1987) Superstring Theory, two volumes. (Cambridge: Cambridge University Press,).
- 42) Guth, A.H. (1981) '*Inflationary universe: A possible solution to the horizon and flatness problems*' *Physical Review D*, **23**, 347-356
- 43) Guth, A.H., (1997) The Inflationary Universe, Reading MA; Addison-Wesley
- 44) Hacking, I., (1983) Representing and Intervening, Cambridge, Cambridge University Press.
- 45) Harvey, J.A., (1999) *Phys. Rev.* **D59**, 026002; [hep-th/9807213].

- 46) Hawking, S., (2001) The Universe in a Nutshell, Bantam Books, NY.
- 47) Hogan, C. J., (1999) "*Why the universe is just so*," astro-ph/9909295.
- 48) Hoyle, F., Burbidge, G., and Narklikar, J.V., COSMOLOGY: The Big Bang's Radical Brother, Cambridge University Press, NY, 2000.
- 49) Hubble, E. (1929) "*A Relation between Distance and Radial Velocity Among Extra-Galactic Nebulae*," Proceedings of the National Academy of Sciences, **15**: 168–173.
- 50) Huey, G., Wang, L., Dave, R., Caldwell, R.R. and Steinhardt, P.J. (1999) "Resolving the Cosmological Missing Energy Problem," *Physical Review D* **59**: 063005, 1–6.
- 51) Huey, G., and Lidsey, J. (2001) Phys. Lett. B 514, 217.
- 52) Jones, S., (1934) General Astronomy, Oxford Press
- 53) Kachru, S., Kumar J., and Silverstein, E., Phys. Rev. **D59**, 106004 (1999) [hep-th/9807076].
- 54) Kawasaki, M, Moroi, T. and Takahashi T., (2001) *Cosmic Microwave Background Anisotropy with Cosine-Type Quintessence*, arXiv: astro-ph/0105161
- 55) Kolb, E.W. (1994) 'Particle Physics and Cosmology', in K.J. Peach and L.L.J. Vick (eds.), High Energy Phenomenology (Fife: Univ. of St. Andrews Press), pp. 361-416
- 56) Kolb, E.W. and Turner, M.S. (1993) *The Early Universe* (Reading: Addison-Wesley)
- 57) Kitcher (1989) "*Explanatory Unification and the Causal Structure of the World*," In: Scientific Explanation, ed. P. Kitcher and W. Salmon, 410-505, University of Minnesota Press, Minneapolis.
- 58) Lemaitre, G. (1927) "*Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques*," Société Scientifique de Bruxelles. Annales A 47: 49–59.
- 59) Lemaitre, G. (1931) "*A Homogeneous Universe of Constant Mass and Increasing Radius Accounting for the Radial Velocity of Extra-galactic Nebulae*," Monthly Notices of the Royal Astronomical Society **91**: 483–489. [English translation of Lemaître (1927).]

- 60) Lemaitre, G. (1933) '*The Uncertainty of the Electromagnetic Field of a Particle*', Physical Review 43, 148
- 61) Lemaitre, G. (1934) '*Evolution of the Expanding Universe*', Proceedings of the National Academy of Sciences of the USA 20, 12-17
- 62) Lemaître, G. (1949) "*The Cosmological Constant*," in Schilpp, P.A. (ed.), Albert Einstein: Philosopher-Scientist, Vol. 2, 439–456. New York: Harper and Row.
- 63) Lewis, David (1991), "Counterfactual Dependence and Time's Arrow", reprinted in Jackson (ed.), *Conditionals*. Oxford: Oxford University Press, 46–75.
- 64) Linde, A. (1974) '*Is the Lee constant a cosmological constant?*' JETP Letters 19, 183-184
- 65) Linde, A.D. (1990) Particle Physics and Inflationary Cosmology (Chur: Harwood Academic Publishers)
- 66) Lykken, J. D., (1996) Introduction to supersymmetry," hep-th/9612114.
- 67) Lykken, J. D., (1999) "*New and improved superstring phenomenology*," astro-ph/9903026.
- 68) Martin, S. P., (1997) A supersymmetry primer," hep-ph/9709356.
- 69) Menzies, P. and Price, H. (1993): "Causation as a Secondary Quality", *British Journal for the Philosophy of Science*, **44**, pp. 187-203.
- 70) Misner, C.W., Thorne, K.S., Wheeler, J.A., (1973), Gravitation, Freeman Press, San Francisco.
- 71) Nernst, W. (1916) "*Über einen Versuch, von quantentheoretischen Betrachtungen zur Annahme stetiger Energiänderungen zurückzukehren*", Verhandlungen der deutsche physikalische Gesellschaft 18, 83-116 43
- 72) Neumann, C. (1896) *Allgemeine Untersuchungen über das Newton'sche Princip der Fernwirkungen mit besonderer Rücksicht auf die elektrische Wirkungen*. Leipzig: Teubner.
- 73) Nilles, H. P. (1984) Phys. Rept. 110, 1.
- 74) Peacock, J.A., (1999) Cosmological Physics, Cambridge, Cambridge University Press.
- 75) Pearl, J. (2000): *Causality*. New York: Cambridge University Press, New York.

- 76) Peebles, P.J.E. and Vilenkin, A. (1999) Phys. Rev. D 59 063505.
- 77) Perlmutter, S. et al. (1999) “*Measurements of Ω and Λ from 42 High Redshift Supernovae*,” Astrophysical Journal **517**: 565–586. Lambda: The Constant That Refuses to Die 219
- 78) Perlmutter, S., Turner, M.S. and White, M. (1999) “*Constraining Dark Energy with Snela and Large-Scale Structure*,” Physical Review Letters **83**: 347–356.
- 79) Petrosian, V. (1974) “*Confrontation of Lemaitre Models and the Cosmological Constant with Observations*,” in Longair, M.S. (ed.), *Confrontation of Cosmological Theories with Observational Data*, 31–46. Dordrecht: D. Reidel.
- 80) Petrosian, V., Salpeter, E. and Sekeres, P. (1967) “*Quasi-Stellar Objects in the Universe with Non-Zero Cosmological Constant*,” Astrophysical Journal **147**: 1222–1226.
- 81) Polchinski, J., (1998) String Theory, two volumes (Cambridge: Cambridge University Press).
- 82) Randall, L., and Sundrum, R. (1999) Phys. Rev. Lett., 83, 3370
- 83) Riess, A.G. et al. (1998) “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant,” *Astronomical Journal* **116**: 1009–1038.
- 84) Rugh, S.E., Zinkernagel, H. (2002) ‘*The quantum vacuum and the cosmological constant problem*’, Studies in History and Philosophy of Modern Physics 33, 663
- 85) Saini et. al (2000) *Reconstructing the Cosmic Equation of State from Supernova Distances*, Phys Rev. Lett., 85, 6, 1162
- 86) Sahni et al. (2003) *Statefinder – a new geometrical diagnostic of dark energy*, arXiv: astro- ph/ 0201498 v2 14 Feb 2003
- 87) Sahni et al. (2002) *The cosmological constant problem and quintessence*, arXiv: astro- ph/ 0202076 v2 9 Mar 2002
- 88) Sahni, V., Sami, M., and Souradeep, T. (2002) Phys. Rev. D 65 023518.
- 89) Salmon, W.C., The Foundations of Scientific Inference, University of Pittsburgh Press, 1967.
- 90) Saunders, S. and Brown, H.R. (1991) ‘*Reflections on Ether*’ in S. Saunders (ed), The Philosophy of vacuum (Oxford: Clarendon Press), 27-63

- 91) Schulmann, R., Kox, A., Janssen, M. and Illy, J. (eds.) (1998) The Collected Papers of Albert Einstein. The Berlin Years: Correspondence, 1914–1918. Princeton, N.J.: Princeton University Press.
- 92) Schmidt, B.G. et al. (1998) “*The High-Z Supernovae Search: Measuring Cosmic Deceleration and Global Curvature of the Universe Using Type Ia Supernovae,*” *Astrophysical Journal* **507**: 46–63.
- 93) Schweber, S.S. (1994) QED and the men who made it: Dyson, Feynman, Schwinger, and Tomonaga, (Princeton, NJ: Princeton University Press)
- 94) J. Schwinger, (1973) ‘*A Report on Quantum Electrodynamics*’, in J. Mehra (ed.), *The Physicist’s Conception of Nature* (Dordrecht: Reidel, 1973), 413 – 429
- 95) Spergel, D.N. and Steinhardt, P.J. (1999) “*Observational Evidence for Self-Interacting Cold Dark Matter,*” astro-ph/9909386.
- 96) Straumann, N. (2002) ‘*The mystery of the cosmic vacuum energy density and the accelerated expansion of the Universe*’ astro-ph/9908342
- 97) Straumann, N. (2004) *The History of the Cosmological Constant Problem,* arXiv:gr-qc/028027 v1
- 98) Turner, M. and White, M. (1997) “CDM Models with a Smooth Component,” *Physical Review D* **56**: 4439–4443.
- 99) Veltman, M. (1975) ‘Cosmology and the Higgs Mass’ *Physical Review Letters* **34**, 777
- 100) Vilenkin, A. (1995), ‘Cosmological Constant and the Evolution of the Universe’ *Phys. Rev. Lett.* **74**, 846, ed. by K. Sato et al. (Universal Academy Press, Tokyo, 1996).
- 101) Vilenkin, A. (2001), “*Cosmological Constant problems and their solutions*”, arXiv:hep-th/0106083
- 102) von Seeliger, H. (1895) “*Über das Newton’sche Gravitationsgesetz,*” *Astronomische Nachrichten*, **137**: 129–136.
- 103) von Seeliger, H. (1896) “*Über das Newton’sche Gravitationsgesetz,*” *Bayerische Akademieder Wissenschaften. Mathematische-Naturwissenschaftliche Klasse* **126**: 373–400.
- 104) von Wright, G.(1971): *Explanation and Understanding.* Ithica, New York: Cornell University Press.

- 105) Wald, R.M. (1994) Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics (Chicago: The University of Chicago Press)
- 106) Wang, L., Caldwell, R.R., Ostriker, J.P. and Steinhardt, P.J. (1999) "*Cosmic Concordance and Quintessence*," astro-ph/9901388.
- 107) Weinberg, S. (1995) The Quantum Theory of Fields. Vol. I. Foundations. Weinberg, S. (1996) The Quantum Theory of Fields. Vol. II. Modern Applications (Cambridge: Cambridge University Press)
- 108) Weinberg, S. (1989) "*The cosmological constant problem*", Review of Modern Physics 61, 1-23
- 109) Weinberg, S. (1997) '*Theories of the Cosmological Constant*' in Turok, N.G.(edt) Critical Dialogues in Cosmology (Singapore: World Scientific), 1-10
- 110) Weinberg, S., (2000) '*The Cosmological Constant Problems*', astro-ph/0005265
- 111) Witten, E., (2000) '*The Cosmological Constant from the viewpoint of string theory*', hep-ph/0002297
- 112) Woodward, J., (2000), "*Explanation and In variance in the Special Sciences*," British Journal of the Philosophy of Science, 51, 197-254
- 113) Zeldovich, Ya.B. (1967) "*Cosmological Constant and Elementary Particles*" Journal of Experimental and Theoretical Physics. Letters **61**: 316–317.
- 114) Zeldovich, Ya.B. (1968) "*The Cosmological Constant and the Theory of Elementary Particles*," Soviet Physics Uspekhi **11**: 381–393.
- 115) Zinkernagel, H. (2002) "*Cosmology, particles and the unity of science*," Studies in Hist and Phil of Mod Phys, 33, 493 – 516
- 116) Zlatev, I. and Steinhardt, P.J. (1999) "*A Tracker Solution to the Cold Dark Matter Cosmic Coincidence Problem*," Physics Letters B **459**: 570–574.
- 117) Zlatev, I., Wang, L. and Steinhardt, P.J. (1999) "*Quintessence, Cosmic Coincidence, and the Cosmological Constant*," Physical Review Letters **82**: 896–899.